

# BREVET D'INVENTION

CERTIFICAT D'UTILITÉ - CERTIFICAT D'ADDITION

## COPIE OFFICIELLE

Le Directeur général de l'Institut national de la propriété industrielle certifie que le document ci-annexé est la copie certifiée conforme d'une demande de titre de propriété industrielle déposée à l'Institut.

Fait à Paris, le 07 DEC. 2004

Pour le Directeur général de l'Institut  
national de la propriété industrielle  
Le Chef du Département des brevets

**PRIORITY DOCUMENT**  
SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH  
RULE 17.1(a) OR (b)

Martine PLANCHE

**BEST AVAILABLE COPY**

INSTITUT  
NATIONAL DE  
LA PROPRIÉTÉ  
INDUSTRIELLE

SIEGE  
26 bis, rue de Saint-Petersbourg  
75800 PARIS cedex 08  
Téléphone : 33 (0)1 53 04 53 04  
Télécopie : 33 (0)1 53 04 45 23  
www.inpi.fr



26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08  
Téléphone : 33 (1) 53 04 53 04 Télécopie : 33 (1) 42 94 86 54

# BREVET D'INVENTION CERTIFICAT D'UTILITÉ

Code de la propriété intellectuelle - Livre VI

cerfa  
N° 11354\*03

## REQUÊTE EN DÉLIVRANCE page 1/2



Cet imprimé est à remplir lisiblement à l'encre noire

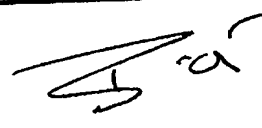
DB 540 08/ 210502

<b>REMISSÉ</b> DATE <b>02 DEC 2003</b> LIEU <b>75 INPI PARIS B</b> N° D'ENREGISTREMENT NATIONAL ATTRIBUÉ PAR L'INPI DATE DE DÉPÔT ATTRIBUÉE PAR L'INPI <b>02 DEC. 2003</b> <b>Vos références pour ce dossier</b> (facultatif) <b>BLOcp226/111FR</b>		<b>1 NOM ET ADRESSE DU DEMANDEUR OU DU MANDATAIRE</b> À QUI LA CORRESPONDANCE DOIT ÊTRE ADRESSÉE  CABINET ORES  36 rue de St Pétersbourg 75008 PARIS	
<b>Confirmation d'un dépôt par télécopie</b>		<input type="checkbox"/> N° attribué par l'INPI à la télécopie	
<b>2 NATURE DE LA DEMANDE</b>		<b>Cochez l'une des 4 cases suivantes</b>	
Demande de brevet		<input checked="" type="checkbox"/>	
Demande de certificat d'utilité		<input type="checkbox"/>	
Demande divisionnaire		<input type="checkbox"/>	
Demande de brevet initiale		N°	Date
ou demande de certificat d'utilité initiale		N°	Date
Transformation d'une demande de brevet européen		<input type="checkbox"/>	
Demande de brevet initiale		N°	Date
<b>3 TITRE DE L'INVENTION</b> (200 caractères ou espaces maximum) UTILISATION DES PROTEINES ET DES PEPTIDES CODES PAR LE GENOME D'UNE NOUVELLE SOUCHE DE CORONAVIRUS ASSOCIE AU SRAS.			
<b>4 DÉCLARATION DE PRIORITÉ</b> <b>OU REQUÊTE DU BÉNÉFICE DE</b> <b>LA DATE DE DÉPÔT D'UNE</b> <b>DEMANDE ANTÉRIEURE FRANÇAISE</b>		Pays ou organisation Date N° Pays ou organisation Date N° Pays ou organisation Date N° <input type="checkbox"/> S'il y a d'autres priorités, cochez la case et utilisez l'imprimé «Suite»	
<b>5 DEMANDEUR</b> (Cochez l'une des 2 cases)		<input checked="" type="checkbox"/> Personne morale <input type="checkbox"/> Personne physique	
Nom ou dénomination sociale		INSTITUT PASTEUR	
Prénoms			
Forme juridique		Etablissement public	
N° SIREN			
Code APE-NAF			
Domicile ou siège	Rue	25-28 rue du Docteur Roux	
	Code postal et ville	75015 PARIS Cedex 15	
	Pays	FRANCE	
Nationalité		Française	
N° de téléphone (facultatif)		N° de télécopie (facultatif)	
Adresse électronique (facultatif)			
<input checked="" type="checkbox"/> S'il y a plus d'un demandeur, cochez la case et utilisez l'imprimé «Suite»			

Remplir impérativement la 2<sup>ème</sup> page

REMISE DES PAGES  
DATE **2 DEC 2003**  
LIEU **75 INPI PARIS B**  
N° D'ENREGISTREMENT  
NATIONAL ATTRIBUÉ PAR L'INPI  
**0314151**

DB 540 W / 210502

<b>6 MANDATAIRE (s'il y a lieu)</b>		RESERVÉ À L'INPI	
Nom		ORES	
Prénom		Béatrice	
Cabinet ou Société		CABINET ORES	
N° de pouvoir permanent et/ou de lien contractuel			
Adresse	Rue	36 rue de St Pétersbourg	
	Code postal et ville	75 010 18 PARIS	
	Pays	FRANCE	
N° de téléphone (facultatif)		01.53.21.11.00.	
N° de télécopie (facultatif)		01.53.21.08.88.	
Adresse électronique (facultatif)		ores@cabinet-ores.com	
<b>7 INVENTEUR (S)</b>		Les inventeurs sont nécessairement des personnes physiques	
Les demandeurs et les inventeurs sont les mêmes personnes		<input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non : Dans ce cas remplir le formulaire de Désignation d'inventeur(s)	
<b>8 RAPPORT DE RECHERCHE</b>		Uniquement pour une demande de brevet (y compris division et transformation)	
Établissement immédiat ou établissement différé		<input checked="" type="checkbox"/> Établissement immédiat <input type="checkbox"/> Établissement différé	
Paiement échelonné de la redevance (en deux versements)		Uniquement pour les personnes physiques effectuant elles-mêmes leur propre dépôt <input type="checkbox"/> Oui <input checked="" type="checkbox"/> Non	
<b>9 RÉDUCTION DU TAUX DES REDEVANCES</b>		Uniquement pour les personnes physiques <input type="checkbox"/> Requête pour la première fois pour cette invention (joindre un avis de non-imposition) <input type="checkbox"/> Obtenue antérieurement à ce dépôt pour cette invention (joindre une copie de la décision d'admission à l'assistance gratuite ou indiquer sa référence) : AG [ ] [ ] [ ] [ ] [ ]	
<b>10 SÉQUENCES DE NUCLEOTIDES ET/OU D'ACIDES AMINÉS</b>		<input checked="" type="checkbox"/> Cochez la case si la description contient une liste de séquences	
Le support électronique de données est joint		<input checked="" type="checkbox"/> Oui <input type="checkbox"/> Non	
La déclaration de conformité de la liste de séquences sur support papier avec le support électronique de données est jointe		<input checked="" type="checkbox"/> Oui <input type="checkbox"/> Non	
Si vous avez utilisé l'imprimé «Suite», indiquez le nombre de pages jointes		1	
<b>11 SIGNATURE DU DEMANDEUR OU DU MANDATAIRE</b> (Nom et qualité du signataire) Le Mandataire, Béatrice ORES (n° 92-4046)		<b>VISA DE LA PRÉFECTURE OU DE L'INPI</b> 	



26 bis, rue de Saint Pétersbourg  
75800 Paris Cedex 08

Téléphone : 33 (1) 53 04 53 04 Télécopie : 33 (1) 42 94 86 54

# BREVET D'INVENTION CERTIFICAT D'UTILITÉ

Code de la propriété Intellectuelle - Livre VI



N° 11354\*03

## REQUÊTE EN DÉLIVRANCE

Page suite N° 1.../1...



Réservé à l'INPI

REMISE DE LA DÉCLARATION

DATE 2 DEC 2003

LIEU 75 INPI PARIS B

0314151

N° D'ENREGISTREMENT

NATIONAL ATTRIBUÉ PAR L'INPI

Cet imprimé est à remplir lisiblement à l'encre noire

DB 829 @ W / 010702

**Vos références pour ce dossier (facultatif)**

BL Ocp226/111FR

**4 DÉCLARATION DE PRIORITÉ  
OU REQUÊTE DU BÉNÉFICE DE  
LA DATE DE DÉPÔT D'UNE  
DEMANDE ANTÉRIEURE FRANÇAISE**

Pays ou organisation

Date

N°

Pays ou organisation

Date

N°

Pays ou organisation

Date

N°

**5 DEMANDEUR (Cochez l'une des 2 cases)**

☒ Personne morale

☐ Personne physique

Nom  
ou dénomination sociale

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Prénoms

Forme juridique

Etablissement public

N° SIREN

Code APE-NAF

Domicile  
ou  
siège

Rue

3 rue Michel-Ange

Code postal et ville

75151 PARIS Cedex 16

Pays

FRANCE

Nationalité

Française

N° de téléphone (facultatif)

N° de télécopie (facultatif)

Adresse électronique (facultatif)

**5 DEMANDEUR (Cochez l'une des 2 cases)**

☐ Personne morale

☐ Personne physique

Nom  
ou dénomination sociale

Prénoms

Forme juridique

N° SIREN

Code APE-NAF

Domicile  
ou  
siège

Rue

Code postal et ville

Pays

Nationalité

N° de téléphone (facultatif)

N° de télécopie (facultatif)

Adresse électronique (facultatif)

**11 SIGNATURE DU DEMANDEUR  
OU DU MANDATAIRE  
(Nom et qualité du signataire)**

Le Mandataire

Béatrice ORES (n° 92-4046)

**VISA DE LA PRÉFECTURE  
OU DE L'INPI**



La présente invention est relative à une nouvelle souche de coronavirus associé au syndrome respiratoire aigu sévère (SRAS), issue d'un prélèvement répertorié sous le n° 031589 et prélevé à Hanoi (Vietnam), à des molécules d'acide nucléique issues de son génome, aux protéines et peptides codés par lesdites  
5 molécules d'acide nucléique ainsi qu'à leurs applications, notamment en tant que réactifs de diagnostic et/ou comme vaccin.

Le coronavirus est un virus à ARN monocaténaire, de polarité positive, d'approximativement 30 kilobases qui se réplique dans le cytoplasme des cellules hôtes ; l'extrémité 5' du génome a une structure en coiffe et l'extrémité 3' comporte  
10 une queue polyA. Ce virus est enveloppé et comprend, à sa surface, -des structures péplomériques- dénommées spicules.

Le génome comprend les cadres ouverts de lecture ou ORF suivants, de son extrémité 5' vers son extrémité 3' : ORF1a et ORF1b correspondant aux protéines du complexe de transcription-réplication, et ORF-S, ORF-E, ORF-M et  
15 ORF-N correspondant aux protéines structurales S, E, M et N. Il comprend également des ORFs correspondant à des protéines de fonction inconnue codées par : la région située entre l'ORF-S et l'ORF-E et chevauchant cette dernière, la région située entre l'ORF-M et l'ORF-N, et la région incluse dans l'ORF-N.

La protéine S est une glycoprotéine membranaire (200-220 kDa) qui  
20 se présente sous la forme de spicules ou "Spike" émergeant de la surface de l'enveloppe virale. Elle est responsable de l'attachement du virus aux récepteurs de la cellule hôte et de l'induction de la fusion de l'enveloppe virale avec la membrane cellulaire.

La petite protéine d'enveloppe (E) également dénommée sM (*small membrane*) qui est une protéine trans-membranaire non glycosylée d'environ 10 kDa,  
25 est la protéine présente en plus faible quantité dans le virion. Elle joue un rôle moteur dans le processus de bourgeonnement des coronavirus qui se produit au niveau du compartiment intermédiaire dans le réticulum endoplasmique et l'appareil de Golgi

La protéine M ou protéine de matrice (25-30 kDa) est une glyco-  
30 protéine membranaire plus abondante qui est intégrée dans la particule virale par une interaction M/E, tandis que l'incorporation de S dans les particules est dirigée par une interaction S/M. Elle semble être importante pour la maturation virale des coronavirus

et pour la détermination du site au niveau duquel les particules virales sont assemblées.

La protéine N ou protéine de nucléocapside (45-50 kDa) qui est la plus conservée parmi les protéines structurales des coronavirus, est nécessaire pour encapsider l'ARN génomique puis pour diriger son incorporation dans le virion. Cette protéine est vraisemblablement également impliquée dans la réplication de l'ARN.

Lorsqu'une cellule hôte est infectée, le cadre de lecture (ORF) situé en 5' du génome viral est traduit en une polyprotéine qui est clivée par les protéases virales et libère alors plusieurs protéines non-structurales telles que l'ARN-polymérase ARN dépendante (Rep) et l'ATPase hélicase (Hel). Ces deux protéines sont impliquées dans la réplication du génome viral ainsi que dans la génération de transcrits qui sont utilisés dans la synthèse des protéines virales. Les mécanismes par lesquels ces ARNms sub-génomiques sont produits, ne sont pas complètement compris ; cependant des faits récents indiquent que les séquences de régulation de la transcription à l'extrémité 5' de chaque gène représentent des signaux qui régulent la transcription discontinue des ARNms sub-génomiques.

Les protéines de la membrane virale (protéines S, E et M) sont insérées dans le compartiment intermédiaire, alors que l'ARN répliqué (brin +) s'assemble avec la protéine N (nucléocapside). Ce complexe protéine-ARN s'associe ensuite avec la protéine M incluse dans les membranes du réticulum endoplasmique et les particules virales se forment lorsque le complexe de la nucléocapside bourgeonne dans le réticulum endoplasmique. Le virus migre ensuite à travers le complexe du Golgi et éventuellement sort de la cellule, par exemple par exocytose. Le site de l'attachement du virus à la cellule hôte se trouve au niveau de la protéine S.

Les coronavirus sont responsables de 15 à 30 % des rhumes chez l'Homme et d'infections respiratoires ou digestives chez les animaux, notamment le chat (FIPV : *Feline infectious peritonitis virus*), la volaille (IBV : *Avian Infectious bronchitis virus*), la souris (MHV : *Mouse Hepatitis virus*), le porc (TGEV : *Transmissible gastroenteritis virus*, PEDV : *Porcine Epidemic Diarrhea virus*, PRCoV : *Porcine Respiratory Coronavirus*, HEV : *Hemagglutinating encephalomyelitis Virus*) et les bovins (BcoV : *Bovine coronavirus*).

En général, chaque coronavirus n'affecte qu'une seule espèce ; chez les individus immunocompétents, l'infection induit des anticorps éventuellement neutralisants et une immunité cellulaire, capables de détruire les cellules infectées.

Une épidémie de pneumonie atypique, dénommée syndrome respiratoire aigu sévère (SARS ou *Severe acute respiratory syndrome*, SRAS en français) s'est propagée dans différents pays (Vietnam, Hong-Kong, Singapour, Thaïlande et Canada) au cours du premier trimestre 2003, à partir d'un foyer initial apparu en Chine dans le dernier trimestre de 2002. La sévérité de cette maladie est telle que son taux de mortalité est d'environ 3 à 6 %. La détermination de l'agent causatif de cette maladie a été entreprise par de nombreux laboratoires, à travers le monde.

En mars 2003, un nouveau coronavirus (SARS-CoV, SARS virus ou virus SRAS, en français) a été isolé, en association avec des cas de syndrome respiratoire aigu sévère (T.G.KSIAZEK et al., *The New England Journal of Medicine*, 2003, 348, 1319-1330 ; C. DROSTEN et al., *The New England Journal of Medicine*, 2003, 348, 1967-1976 ; Peiris et al., *Lancet*, 2003, 361, 1319-).

Des séquences génomiques de ce nouveau coronavirus ont ainsi été obtenues, notamment celles de l'isolat Urbani (Genbank n° d'accès AY274119.3 et A. MARRA et al., *Science*, May 1, 2003, 300, 1399-1404) et de l'isolat de Toronto (Tor2, Genbank n° d'accès AY 278741 et A. ROTA et al., *Science*, 2003, 300, 1394-1399).

L'organisation du génome est comparable à celle des autres coronavirus connus permettant ainsi de confirmer l'appartenance du SARS-CoV à la famille des *Coronaviridae* ; les cadres ouverts de lecture ORF1a et 1b et les cadres ouverts de lecture correspondant aux protéines S, E, M, et N, ainsi qu'à des protéines codées par : la région située entre l'ORF-S et l'ORF-E (ORF3), la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E (ORF4), la région située entre l'ORF-M et l'ORF-N (ORF7 à ORF11) et la région correspondant à l'ORF-N (ORF13 et ORF14), ont notamment été identifiées.

Sept différences ont été mises en évidence entre les séquences des isolats Tor2 et Urbani ; 3 correspondent à des mutations silencieuses (c/t en position 16622 et a/g en position 19064 de l'ORF1b, t/c en position 24872 de l'ORF-S) et 4 modifient la séquence en acides aminés de respectivement : les protéines codées par

l'ORF1a (c/t en position 7919 correspondant à la mutation A/V), la protéine S (g/t en position 23220 correspondant à la mutation A/S), la protéine codée par l'ORF3 (a/g en position 25298 correspondant à la mutation R/G) et de la protéine M (t/c en position 26857 correspondant à la mutation S/P).

5 En outre, l'analyse phylogénétique montre que le SARS-CoV est éloigné des autres coronavirus et qu'il est apparu, ni par mutation de coronavirus respiratoires humains, ni par recombinaison entre des coronavirus connus (pour une revue, voir Holmes, J.C.I., 2003, 111, 1605-1609).

10 La mise en évidence et la prise en compte de nouveaux variants sont importantes pour la mise au point de réactifs de détection et de diagnostic du SRAS suffisamment sensibles et spécifiques ainsi qu'à des compositions immunogènes aptes à protéger des populations contre des épidémies de SRAS.

Les Inventeurs ont maintenant mis en évidence une autre souche de coronavirus associé au SRAS, qui se distingue des isolats Tor2 et Urbani.

15 La présente invention a donc pour objet, une souche isolée ou purifiée de coronavirus humain associé au syndrome respiratoire aigu sévère, caractérisée en ce que son génome présente sous la forme d'ADN complémentaire un codon sérine en position 23220-23222 du gène de la protéine S ou un codon glycine en position 25298-25300 du gène de l'ORF3, et un codon alanine en position 7918-7920 de  
20 l'ORF1a ou un codon sérine en position 26857-26859 du gène de la protéine M, lesdites positions étant indiquées en référence à la séquence Genbank AY274119.3.

Selon un mode de réalisation avantageux de ladite souche, l'équivalent ADN de son génome présente une séquence correspondant à la séquence SEQ ID NO : 1 ; cette souche de coronavirus est issue du prélèvement de lavage  
25 bronchoalvéolaire d'un patient atteint de SRAS, répertorié sous le n° 031589 et effectué à l'hôpital français de Hanoi (Vietnam).

Conformément à l'invention, ladite séquence SEQ ID NO :1 est celle de l'acide désoxyribonucléique correspondant à la molécule d'acide ribonucléique du génome de la souche isolée de coronavirus telle que définie ci-  
30 dessus.

La séquence SEQ ID NO : 1 se distingue de la séquence Genbank AY274119.3 (isolat Tor2) en ce qu'elle possède les mutations suivantes :

- g/t en position 23220 ; le codon alanine (gct) en position 577 de la séquence en acides aminés de la protéine S de Tor2 est remplacé par un codon sérine (tct),

5 - a/g en position 25298 ; le codon arginine (aga) en position 11 de la séquence en acide aminés de la protéine codée par l'ORF3 de Tor 2 est remplacé par un codon glycine (gga).

En outre, la séquence SEQ ID NO : 1 se distingue de la séquence Genbank AY278741 (isolat Urbani) en ce qu'elle possède les mutations suivantes :

10 - t/c en position 7919 ; le codon valine (ggt) en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a est remplacé par un codon alanine (gct),

- t/c en position 16622 : cette mutation ne modifie pas la séquence en acides aminés des protéines codées par l'ORF1b (mutation silencieuse),

15 - g/a en position 19064 : cette mutation ne modifie pas la séquence en acides aminés des protéines codées par l'ORF1b (mutation silencieuse),

- c/t en position 24872 : cette mutation ne modifie pas la séquence en acides aminés de la protéine S, et

- c/t en position 26857 : le codon proline (ccc) en position 154 de la séquence en acides aminés de la protéine M est remplacé par un codon sérine (tcc).

20 En l'absence de mention particulière, les positions des séquences nucléotidiques et peptidiques sont indiquées en référence à la séquence Genbank AY274119.3.

La présente invention a également pour objet un polynucléotide isolé ou purifié, caractérisé en ce que sa séquence est celle du génome de la souche isolée  
25 de coronavirus telle que définie ci-dessus.

Selon un mode de réalisation avantageux dudit polynucléotide il présente la séquence SEQ ID NO : 1.

La présente invention a également pour objet un polynucléotide isolé ou purifié, caractérisé en ce que sa séquence hybride dans des conditions de forte  
30 stringence avec la séquence du polynucléotide tel que défini ci-dessus.

Les termes « isolé ou purifié » signifient modifié « par la main de l'homme » à partir de l'état naturel ; autrement dit si un objet existe dans la nature, il

est dit isolé ou purifié s'il a été modifié ou extrait de son environnement naturel ou les deux. Par exemple, un polynucléotide ou une protéine/un peptide naturellement présent dans un organisme vivant n'est ni isolé, ni purifié ; en revanche le même polynucléotide ou protéine /peptide séparé des molécules coexistantes dans son environnement naturel, obtenu par clonage, amplification et/ou synthèse chimique est isolé au sens de la présente invention. De plus, un polynucléotide ou une protéine/peptide qui est introduit dans un organisme par transformation, manipulation génétique ou par toute autre méthode, est « isolé » même s'il est présent dans ledit organisme. Le terme purifié tel qu'utilisé dans la présente invention, signifie que les protéines /peptides selon l'invention sont essentiellement libres d'association avec les autres protéines ou polypeptides, comme l'est par exemple le produit purifié de la culture de cellules hôtes recombinantes ou le produit purifié à partir d'une source non-recombinante.

Au sens de la présente invention, on entend par conditions d'hybridation de forte stringence, des conditions de température et de force ionique choisies de telle manière qu'elles permettent le maintien de l'hybridation spécifique et sélective entre polynucléotides complémentaires.

A titre d'illustration, des conditions de forte stringence aux fins de définir les polynucléotides ci-dessus, sont avantageusement les suivantes : l'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M citrate de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation pendant 20 heures à 42°C suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 20°C en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS pendant 30 minutes à 60°C.

La présente invention a également pour objet un fragment représentatif du polynucléotide tel que défini ci-dessus, caractérisé en ce qu'il est susceptible d'être obtenu, soit par l'utilisation d'enzymes de restriction dont les sites de reconnaissance et de coupure sont présents dans ledit polynucléotide tel que défini ci-dessus, soit par amplification à l'aide d'amorces oligonucléotidiques spécifiques dudit polynucléotide tel que défini ci-dessus, soit par transcription *in vitro*, soit par synthèse

chimique.

Selon un mode de réalisation avantageux dudit fragment, il est sélectionné dans le groupe constitué par : l'ADNc correspondant à au moins un cadre ouvert de lecture (ORF) choisi parmi : ORF1a, ORF1b, ORF-S, ORF-E, ORF-M, ORF-N, ORF3, ORF4, ORF7 à ORF11, ORF13 et ORF14, et l'ADNc correspondant aux extrémités 5' ou 3' non-codantes dudit polynucléotide.

Selon une disposition avantageuse de ce mode de réalisation, ledit fragment présente une séquence sélectionnée dans le groupe constitué par :

- les séquences SEQ ID NO : 2 et 4 représentant l'ADNc correspondant à l'ORF-S qui code pour la protéine S,
- les séquences SEQ ID NO : 13 et 15 représentant l'ADNc correspondant à l'ORF-E qui code pour la protéine E,
- les séquences séquence SEQ ID NO : 16 et 18 représentant l'ADNc correspondant à l'ORF-M qui code pour la protéine M,
- les séquences SEQ ID NO : 36 et 38 représentant l'ADNc correspondant à l'ORF-N qui code pour la protéine N,
- les séquences représentant les ADNc correspondant respectivement : aux ORF1a et ORF1b (ORF1ab, SEQ ID NO : 31), aux ORF3 et ORF4 (SEQ ID NO : 7, 8), aux ORF 7 à 11 (SEQ ID NO : 19, 20), à l'ORF13 (SEQ ID NO : 32) et à l'ORF14 (SEQ ID NO : 34), et
- les séquences représentant les ADNc correspondant respectivement aux extrémités 5'(SEQ ID NO : 39 et 72) et 3' non-codantes (SEQ ID NO : 40, 73) dudit polynucléotide.

La présente invention a également pour objet un fragment de l'ADNc codant pour la protéine S, tel que défini ci-dessus, caractérisé en ce qu'il présente une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 5 et 6 (fragments Sa et Sb).

La présente invention a également pour objet un fragment de l'ADNc correspondant aux ORF1a et ORF1b tel que défini ci-dessus, caractérisé en ce qu'il présente une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 41 à 54 (fragments L0 à L12).

La présente invention a également pour objet un fragment du polynuéotide tel que défini ci dessus, caractérisé en ce qu'il présente au moins 15 bases ou paires de bases consécutives de la séquence du génome de ladite souche incluant au moins une de celles situées en position 7979, 16622, 19064, 23220, 24872, 25298 et 5 26857. De préférence, il s'agit d'un fragment de 20 à 2500 bases ou paires de bases, de manière préférée de 20 à 400.

Selon un mode de réalisation avantageux dudit fragment, il inclut au moins un couple de bases ou de paires de bases correspondant aux positions suivantes : 7919 et 23220, 7919 et 25298, 16622 et 23220, 19064 et 23220, 16622 et 10 25298, 19064 et 25298, 23220 et 24872, 23220 et 26857, 24872 et 25298, 25298 et 26857.

La présente invention a également pour objet des amorces d'au moins 18 bases aptes à amplifier un fragment du génome d'un coronavirus associé au SRAS ou de l'équivalent ADN de celui-ci.

15 Selon un mode de réalisation desdites amorces, elles sont sélectionnées dans le groupe constitué par :

- la paire d'amorces n° 1 correspondant respectivement aux positions 28507 à 28522 (amorce sens, SEQ ID NO : 60) et 28774 à 28759 (amorce anti-sens, SEQ ID NO : 61) de la séquence du polynuéotide tel que défini ci-dessus, et

20 - la paire d'amorces n° 2 correspondant respectivement aux positions 28375 à 28390 (amorce sens, SEQ ID NO : 62) et 28702 à 28687 (amorce anti-sens, SEQ ID NO : 63) de la séquence du polynuéotide tel que défini ci-dessus.

La présente invention a également pour objet une sonde apte à détecter la présence du génome d'un coronavirus associé au SRAS ou d'un fragment 25 de celui-ci, caractérisée en ce qu'elle est sélectionnée dans le groupe constitué par : les fragments tels que définis ci-dessus et les fragments correspondant aux positions suivantes de la séquence du polynuéotide tel que défini ci-dessus : 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 (SEQ ID NO : 64 à 67).

Les sondes et amorces selon l'invention peuvent être marquées 30 directement ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'Homme du Métier, afin d'obtenir un signal détectable et/ou quantifiable. Parmi les isotopes radioactifs utilisés, on peut citer le <sup>32</sup>P, le <sup>33</sup>P, le



$^{35}\text{S}$ , le  $^3\text{H}$  ou l' $^{125}\text{I}$ . Les entités non radioactives sont sélectionnées parmi les ligands tels que la biotine, l'avidine, la streptavidine, la digoxygénine, les haptènes, les colorants, les agents luminescents tels que les agents radioluminescents, chémoluminescents, bioluminescents, fluorescents, phosphorescents.

5 L'invention englobe les sondes et les amorces marquées dérivées des séquences précédentes.

De telles sondes et amorces sont utiles pour le diagnostic de l'infection par un coronavirus associé au SRAS.

La présente invention a également pour objet une méthode de détection  
10 tion d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

(a) l'extraction d'acides nucléiques présents dans ledit échantillon biologique,

(b) l'amplification d'un fragment de l'ORF-N par RT-PCR à l'aide  
15 d'une paire d'amorces telle que définie ci-dessus, et

(c) la détection par tout moyen approprié des produits d'amplifications obtenus en (b).

Les produits d'amplifications (amplicons) en (b) sont de 268 pb pour la paire d'amorces n° 1 et de 328 pb pour la paire d'amorces n°2.

20 Selon un mode de mise en œuvre avantageux dudit procédé, l'étape (b) de détection est réalisée à l'aide d'au moins une sonde correspondant aux positions 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 de la séquence du polynucléotide tel que défini ci-dessus.

De préférence, le génome du coronavirus associé au SRAS est  
25 détecté et éventuellement quantifié par PCR, en temps réel, à l'aide de la paire d'amorces n°2 et des sondes correspondant aux positions 28541 à 28563 et 28565 à 28589 marquées avec des composés différents, notamment des agents fluorescents différents.

La RT-PCR en temps réel qui met en œuvre cette paire d'amorces et  
30 cette sonde est très sensible puisqu'elle permet de détecter  $10^2$  copies d'ARN et jusqu'à 10 copies d'ARN, elle est en outre fiable et reproductible.

L'invention englobe les polydésoxyribonucléotides et les polyribonucléotides simple-brin, double-brin et tripe-brin correspondant à la séquence du génome de la souche isolée de coronavirus et de ses fragments tels que définis ci-dessus, ainsi qu'à leurs séquences complémentaires, sens ou anti-sens, notamment les  
 5 ARN et les ADNc correspondant à la séquence du génome et de ses fragments tels que définis ci-dessus.

La présente invention englobe également les fragments d'amplification obtenus à l'aide d'amorces spécifiques du génome de la souche purifiée ou isolée tel que défini ci-dessus, notamment à l'aide d'amorces et de paires  
 10 d'amorces telles que définies ci-dessus, les fragments de restriction constitués par ou comprenant la séquence des fragments tels que définis ci-dessus, les fragments obtenus par transcription *in vitro* à partir d'un vecteur contenant la séquence SEQ ID NO : 1 ou un fragment tel que défini ci-dessus, ainsi que des fragments obtenus par synthèse chimique. Des exemples de fragments de restriction sont déduits de la carte  
 15 de restriction de la séquence SEQ ID NO : 1 illustrée par la figure 13. Conformément à l'invention lesdits fragments sont, soit sous forme de fragments isolés, soit sous forme de mélanges de fragments. L'invention englobe également les fragments modifiés, par rapport aux précédents, par enlèvement, ou addition de nucléotides dans une proportion d'environ 15 %, par rapport à la longueur des fragments ci-dessus et/ou  
 20 modifiés au niveau de la nature des nucléotides, dès lors que les fragments nucléotidiques modifiés conservent une capacité d'hybridation avec les séquences d'ARN génomiques ou antigénomiques de l'isolat tel que défini ci-dessus.

Les molécules d'acide nucléique selon l'invention sont obtenues par les méthodes classiques, connues en elles-mêmes, en suivant les protocoles standards  
 25 tels que ceux décrits dans *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA). Par exemple, elles peuvent être obtenues par amplification d'une séquence nucléique par PCR ou RT-PCR ou bien par synthèse chimique totale ou partielle.

La présente invention a également pour objet une puce ou filtre à  
 30 ADN ou à ARN, caractérisé en ce qu'il comprend au moins un polynucléotide ou l'un de ses fragments tels que définis ci-dessus.

Les puces ou filtres à ADN ou à ARN selon l'invention sont

préparés par les méthodes classiques, connues en elles-mêmes, comme par exemple greffage chimique ou électrochimique d'oligonucléotides sur support de verre ou de nylon.

La présente invention a également pour objet un vecteur de clonage et/ou d'expression recombinant, notamment un plasmide ou un phage comprenant un fragment d'acide nucléique tel que défini ci-dessus. De préférence, ledit vecteur recombinant est un vecteur d'expression dans lequel ledit fragment d'acide nucléique est placé sous le contrôle d'éléments régulateurs de la transcription et de la traduction appropriés. En outre, ledit vecteur peut comprendre des séquences (étiquettes ou *tag*) fusionnées en phase avec l'extrémité 5' et/ou 3' dudit insert, utiles pour l'immobilisation, et/ou la détection et/ou la purification de la protéine exprimée à partir dudit vecteur.

Ces vecteurs sont construits et introduits dans des cellules hôtes par les méthodes classiques d'ADN recombinant et de génie génétique, qui sont connues en elles-mêmes. De nombreux vecteurs dans lesquels on peut insérer une molécule d'acide nucléique d'intérêt afin de l'introduire et de la maintenir dans une cellule hôte, sont connus en eux-mêmes ; le choix d'un vecteur approprié dépend de l'utilisation envisagée pour ce vecteur (par exemple répllication de la séquence d'intérêt, expression de cette séquence, maintien de la séquence sous forme extrachromosomique ou bien intégration dans le matériel chromosomique de l'hôte), ainsi que de la nature de la cellule hôte.

Conformément à l'invention, ledit plasmide est notamment sélectionné parmi les plasmides suivants :

- le plasmide, dénommé SARS-S, compris dans la souche bactérienne déposée sous le n° I-3059, le 20 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, laquelle séquence correspondant aux nucléotides des positions 21406 à 25348 (SEQ ID NO : 4), en référence à la séquence Genbank AY274119.3,

- le plasmide, dénommé SARS-S1, compris dans la souche bactérienne déposée sous le n° I-3020, le 12 mai 2003, auprès de la Collection Nationale de

Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 5' de la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment correspondant aux nucléotides des positions 21406 à 23454 (SEQ ID NO :5), en référence à la séquence Genbank AY274119.3 Tor2,

- le plasmide, dénommé SARS-S2, compris dans la souche bactérienne déposée sous le n° I-3019, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 3' de la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment correspondant aux nucléotides des positions 23322 à 25348 (SEQ ID NO :6), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide, dénommé SARS-SE, compris dans la souche bactérienne déposée sous le n° I-3126, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15; il contient l'ADNc correspondant à la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle région correspondant aux nucléotides des positions 25110 à 26244 (SEQ ID NO :8), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide, dénommé SARS-E, compris dans la souche bactérienne déposée sous le n° I-3046, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 26082 à 26413 (SEQ ID NO :15), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide, dénommé SARS-M ; compris dans la souche bactérienne déposée sous le n° I-3047, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus ;

laquelle séquence correspondant aux nucléotides des positions 26330 à 27098 (SEQ ID NO :18), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-MN, compris dans la souche bactérienne déposée sous le n° I-3125, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc correspondant à la région située entre l'ORF-M et l'ORF-N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 et prélevée à Hanoi, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 26977 à 28218 (SEQ ID NO :20), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-N, compris dans la souche bactérienne déposée sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc codant pour la protéine N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 28054 à 29430 (SEQ ID NO :38), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-5'NC, compris dans la souche bactérienne déposée sous le n° I- 3124, le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc correspondant à l'extrémité 5'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 1 à 204 (SEQ ID NO :39), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-3'NC, compris dans la souche bactérienne déposée sous le n° I-3123 le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15. ; il contient la séquence d'ADNc correspondant à l'extrémité 3'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant à celle située entre le nucléotide en position 28933 à 29727 (SEQ ID NO :40), en référence à la séquence Genbank n° d'accès AY274119.3, se termine par une série de nucléotides a.,

- le plasmide d'expression dénommé pIV2.3N, contenant un fragment d'ADNc codant pour une fusion C-terminale de la protéine N (SEQ ID NO : 37) avec une étiquette polyhistidine,

5 - le plasmide d'expression dénommé pIV2.3S<sub>C</sub>, contenant un fragment d'ADNc codant pour une fusion C-terminale du fragment correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine,

10 - le plasmide d'expression pIV2.3S<sub>L</sub>, contenant un fragment d'ADNc codant pour une fusion C-terminale du fragment correspondant aux positions 14 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine,

- le plasmide d'expression dénommé pIV2.4N, contenant un fragment d'ADNc codant pour une fusion N-terminale de la protéine N (SEQ ID NO : 3) avec une étiquette polyhistidine,

15 - le plasmide d'expression dénommé pIV2.4S<sub>C</sub> ou pIV2.4S<sub>1</sub>, contenant un insert codant pour une fusion N-terminale du fragment correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine, et

20 - le plasmide d'expression dénommé pIV2.4S<sub>L</sub> contenant un fragment d'ADNc codant pour une fusion N-terminale du fragment correspondant aux positions 14 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine.

Selon une disposition avantageuse du plasmide d'expression tel que défini ci-dessus, il est compris dans une souche bactérienne qui a été déposée sous le  
25 n° I- 3117, le 23 octobre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

Selon une autre disposition avantageuse du plasmide d'expression tel que défini ci-dessus, il est compris dans une souche bactérienne qui a été déposée sous le n° I- 3118, le 23 octobre 2003, auprès de la Collection Nationale de Cultures  
30 de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

La présente invention a également pour objet une banque d'ADNc caractérisée en ce qu'elle comprend des fragments tels que définis ci-dessus, en parti-

culier des fragments d'amplification ou des fragments de restriction, clonés dans un vecteur recombinant, notamment un vecteur d'expression (banque d'expression).

La présente invention a également pour objet des cellules, notamment des cellules procaryotes, modifiées par un vecteur recombinant tel que défini ci-dessus.

Les vecteurs recombinants tels que définis ci-dessus et les cellules transformées par lesdits vecteurs d'expression sont avantageusement utilisés pour la production des protéines et des peptides correspondants. Les banques d'expression dérivées desdits vecteurs, ainsi que les cellules transformées par lesdites banques d'expression sont avantageusement utilisées pour identifier les épitopes immunogènes (épitopes B et T) des protéines du coronavirus associé au SRAS.

La présente invention a également pour objet les protéines et les peptides purifiées ou isolées, caractérisés en ce qu'ils sont codés par le polynucléotide ou l'un de ses fragments tels que définis ci-dessus.

Selon un mode de réalisation avantageux de l'invention, ladite protéine est sélectionnée dans le groupe constitué par :

- la protéine S de séquence SEQ ID NO :3
- la protéine E de séquence SEQ ID NO :14
- la protéine M de séquence SEQ ID NO :17
- la protéine N de séquence SEQ ID NO : 37
- les protéines codées par les ORFs : ORF1a, ORF1b, ORF3, ORF4 et ORF7 à ORF11, ORF13 et ORF14 de séquence respectivement, SEQ ID NO :74, 75, 10, 12, 22, 24, 26, 28, 30, 33 et 35.

Selon un mode de réalisation avantageux de l'invention, ledit peptide est sélectionné dans le groupe constitué par :

- a) les peptides correspondant aux positions 14 à 1193 et 475 à 1193 de la séquence en acides aminés de la protéine S,
- b) les peptides correspondant aux positions 2 à 14 (SEQ ID NO : 69) et 100 à 221 de la séquence en acides aminés de la protéine M ; ces peptides correspondent respectivement à l'ectodomaine et à l'endodomaine de la protéine M, et
- c) les peptides correspondant aux positions 1 à 12 (SEQ ID NO : 70) et 53 à 76 (SEQ ID NO : 71) de la séquence en acides aminés de la protéine E ; ces

peptides correspondent respectivement à l'ectodomaine et à l'extrémité C-terminale de la protéine E, et

- d) les peptides de 5 à 50 acides aminés consécutifs, de préférence de 10 à 30 acides aminés, inclus ou chevauchant partiellement ou totalement la séquence des peptides tels que définis en a), b) ou c).

La présente invention a également pour objet un peptide caractérisé en ce qu'il présente une séquence de 7 à 50 acides aminés incluant un résidu d'acide aminé sélectionné dans le groupe constitué par :

- l'alanine située en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a.
- la sérine située en position 577 de la séquence en acides aminés de la protéine S de la souche de SARS-CoV telle que définie ci-dessus,
- la glycine en position 11 de la séquence en acides aminés de la protéine codée par l'ORF3 de la souche de SARS-CoV telle que définie ci-dessus,
- la sérine en position 154 de la séquence en acides aminés de la protéine M de la souche de SARS-CoV telle que définie ci-dessus.

- La présente invention a également pour objet un anticorps ou un fragment d'anticorps polyclonal ou monoclonal, susceptible d'être obtenu par immunisation d'un animal avec un vecteur recombinant tel que défini ci-dessus, une banque d'ADNc telle que définie ci-dessus ou bien une protéine ou un peptide tels que définis ci-dessus, caractérisé en ce qu'il se lie avec l'une au moins des protéines codées par le SARS-CoV telles que définies ci-dessus.

- L'invention englobe les anticorps polyclonaux, les anticorps monoclonaux, les anticorps chimériques tels que les anticorps humanisés, ainsi que leurs fragments (Fab, Fv, scFv).

- Au sens de la présente invention, on entend par anticorps chimérique, relativement à un anticorps d'une espèce animale particulière ou d'une classe particulière d'anticorps, un anticorps comprenant tout ou partie d'une chaîne lourde et/ou d'une chaîne légère d'un anticorps d'une autre espèce animale ou d'une autre classe d'anticorps.

Au sens de la présente invention, on entend par anticorps humanisé une immunoglobuline humaine dans laquelle les résidus des CDRs



(*Complementary-Determining Regions*) qui forment le site de liaison à l'antigène sont remplacés par ceux d'un anticorps monoclonal non-humain possédant la spécificité, l'affinité ou l'activité recherchées. Par comparaison avec les anticorps non-humains, les anticorps humanisés sont moins immunogènes et possèdent une demi-vie prolongée chez l'Homme car ils ne possèdent qu'une faible proportion de séquences non-humaines étant donné que la quasi-totalité des résidus des régions FR (Framework) et de la région constante (Fc) de ces anticorps sont ceux d'une séquence consensus d'immunoglobulines humaines.

La présente invention a également pour objet une puce à protéine, caractérisée en ce qu'elle comprend une protéine, un peptide ou bien un anticorps tels que définis ci-dessus.

Les puces à protéine selon l'invention sont préparées par les méthodes classiques, connues en elles-mêmes. Parmi les supports appropriés sur lesquels peuvent être immobilisés des protéines, on peut citer ceux en matière plastique ou en verre, notamment sous la forme de microplaques.

La présente invention a également pour objet des réactifs dérivés de la souche isolée de coronavirus associé au SRAS, issue du prélèvement répertorié sous le n° 031589, utiles pour l'étude et le diagnostic de l'infection provoquée par un coronavirus associé au SRAS, lesquels réactifs sont sélectionnés dans le groupe constitué par :

- (a) une paire d'amorces, une sonde ou une puce à ADN telles que définies ci-dessus,
- (b) un vecteur recombinant ou une cellule modifiée tels que définis ci-dessus,
- (c) une souche isolée de coronavirus ou un polynucléotide tels que définis ci-dessus,
- (d) une protéine ou un peptide tel que défini ci-dessus,
- (e) un anticorps ou fragment d'anticorps tels que définis ci-dessus, et
- (f) une puce à protéine telle que définie ci-dessus.

Ces différents réactifs sont préparés et utilisés selon les techniques classiques de biologie moléculaire et d'immunologie, en suivant les protocoles standards tels que ceux décrits dans *Current Protocols in Molecular Biology*

(Frederick M. AUSUBEL, 2000, Wiley and Son Inc., Library of Congress, USA), dans *Current Protocols in Immunology* (John E. Cologan, 2000, Wiley and Son Inc. Library of Congress, USA) et dans *Antibodies : A Laboratory Manual* (E. Howell and D Lane, Cold Spring Harbor Laboratory, 1988).

5 Les fragments d'acide nucléique selon l'invention sont préparés et utilisés selon les techniques classiques telles que définies ci-dessus. Les peptides et les protéines selon l'invention sont préparés par les techniques d'ADN recombinant, connues de l'Homme du métier, notamment à l'aide des vecteurs recombinants tels que définis ci-dessus. Alternativement, les peptides selon l'invention peuvent être  
10 préparés par les techniques classiques de synthèse en phase solide ou liquide, connues de l'Homme du métier.

Les anticorps polyclonaux sont préparés par immunisation d'un animal approprié avec une protéine ou un peptide tels que définis ci-dessus, éventuellement couplé à la KLH ou à l'albumine et/ou associé à un adjuvant approprié tel que  
15 l'adjuvant de Freund (complet ou incomplet) ou l'hydroxyde d'alumine ; après obtention d'un titre en anticorps satisfaisant, les anticorps sont récoltés par prélèvement du sérum des animaux immunisés et enrichis en IgG par précipitation, selon les techniques classiques, puis les IgG spécifiques des protéines du SARS-CoV sont éventuellement purifiées par chromatographie d'affinité sur une colonne appropriée  
20 sur laquelle sont fixés ledit peptide ou ladite protéine, tels que définis ci-dessus, de façon à obtenir une préparation d'IgG monospécifiques.

Les anticorps monoclonaux sont produits à partir d'hybridomes obtenus par fusion de lymphocytes B d'un animal immunisé par une protéine ou un peptide tels que définis ci-dessus avec des myélomes, selon la technique de Köhler et  
25 Milstein (Nature, 1975, 256, 495-497) ; les hybridomes sont cultivés *in vitro*, notamment dans des fermenteurs ou produits *in vivo*, sous forme d'ascite ; alternativement lesdits anticorps monoclonaux sont produits par génie génétique comme décrit dans le brevet américain US 4,816,567.

Les anticorps humanisés sont produits par des méthodes générales  
30 comme celles décrites dans la Demande Internationale WO 98/45332.

Les fragments d'anticorps sont produits à partir des régions V<sub>H</sub> et V<sub>L</sub> clonées, à partir des ARNm d'hybridomes ou de lymphocytes spléniques d'une

souris immunisée ; par exemple, les fragments Fv, scFv ou Fab sont exprimés à la surface de phages filamenteux selon la technique de Winter et Milstein (Nature, 1991, 349, 293-299) ; après plusieurs étapes de sélection, les fragments d'anticorps spécifiques de l'antigène sont isolés et exprimés dans un système d'expression approprié, par les techniques classiques de clonage et d'expression d'ADN recombinant.

Les anticorps ou leur fragments tels que définis ci-dessus, sont purifiés par les techniques classiques connues de l'Homme du métier, telles que la chromatographie d'affinité.

La présente invention a en outre pour objet l'utilisation d'un produit sélectionné dans le groupe constitué par : une paire d'amorces, une sonde, une puce à ADN, un vecteur recombinant, une cellule modifiée, une souche isolée de coronavirus, un polynucléotide, une protéine ou un peptide, un anticorps ou un fragment d'anticorps, et une puce à protéine tels que définis ci-dessus, pour la préparation d'un réactif de détection et éventuellement de génotypage/sérotypage, d'un coronavirus associé au SRAS.

Les protéines et les peptides selon l'invention, qui sont aptes à être reconnus et/ou à induire la production d'anticorps spécifiques du coronavirus associé au SRAS, sont utiles pour le diagnostic de l'infection par un tel coronavirus ; l'infection est détectée, par une technique appropriée- notamment EIA, ELISA, RIA, immunofluorescence-, à partir d'un échantillon biologique prélevé chez un individu susceptible d'être infecté.

Selon une disposition avantageuse de ladite utilisation, lesdites protéines sont sélectionnées dans le groupe constitué par les protéines S, E, M et/ou N et les peptides tels que définis ci-dessus.

Les protéines S, E, M et/ou N et les peptides dérivés de ces protéines tels que définis ci-dessus, par exemple la protéine N, sont utilisées pour le diagnostic indirect d'une infection à coronavirus associé au SRAS (diagnostic sérologique ; détection d'anticorps spécifiques du SARS-CoV), notamment par une méthode immunoenzymatique (ELISA).

Les anticorps et les fragments d'anticorps selon l'invention, notamment ceux dirigés contre les protéines S, E, M et/ou N et les peptides dérivés tels que définis ci-dessus, sont utiles pour le diagnostic direct d'une infection à coro-

navirus associé au SRAS ; la détection de protéine(s) du SARS-CoV est réalisée par une technique appropriée, notamment EIA, ELISA, RIA, immunofluorescence à partir d'un échantillon biologique prélevé chez un individu susceptible d'être infecté.

La présente invention a également pour objet une méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

- (a) la mise en contact dudit échantillon biologique avec au moins un anticorps ou un fragment d'anticorps, une protéine, un peptide ou bien une puce ou un filtre à protéine ou à peptide tels que définis ci-dessus, et
- (b) la révélation par tout moyen approprié des complexes antigène-anticorps formés en (a), par exemple par EIA, ELISA, RIA, ou par immunofluorescence.

Selon un mode de mise en œuvre avantageux dudit procédé l'étape

- (a) comprend :
  - (a<sub>1</sub>) la mise en contact dudit échantillon biologique avec au moins un premier anticorps ou un fragment d'anticorps qui est fixé sur un support approprié, notamment une microplaque,
  - (a<sub>2</sub>) le lavage de la phase solide, et
  - (a<sub>3</sub>) l'addition d'au moins un second anticorps ou un fragment d'anticorps, différent du premier, ledit anticorps ou fragment d'anticorps étant éventuellement marqué de façon appropriée.

Ce procédé qui permet de capturer les particules virales présentes dans l'échantillon biologique est également dénommé procédé d'immunocapture.

Par exemple :

- l'étape (a<sub>1</sub>) est réalisée avec au moins un premier anticorps monoclonal ou polyclonal ou un fragment de ceux-ci, dirigé contre la protéine S, M, et/ou E, et/ou un peptide correspondant à l'ectodomaine de l'une de ces protéines (peptides M2-14 ou E1-12)
- l'étape (a<sub>3</sub>) est réalisée avec au moins un anticorps ou un fragment d'anticorps dirigé contre un autre épitope de la même protéine ou de préférence contre une autre protéine, de manière préférée contre une protéine interne telle que la nucléoprotéine N ou l'endodomaine de la protéine E ou M, de manière encore plus préférée

il s'agit d'anticorps ou de fragments d'anticorps dirigés contre la protéine N qui est très abondante dans la particule virale ; lorsqu'un anticorps ou un fragment d'anticorps dirigé contre une protéine interne (N) ou contre l'endodomaine des protéines E ou M est utilisé, le dit anticorps est incubé en présence de détergent, comme le Tween 20 par exemple, à des concentrations de l'ordre de 0,1 %.

- l'étape (b) de révélation des complexes antigène-anticorps formés est réalisée, soit directement à l'aide d'un second anticorps marqué par exemple avec de la biotine ou une enzyme appropriée telle que la peroxydase ou la phosphatase alcaline, soit indirectement à l'aide d'un sérum anti-immunoglobulines marqué comme ci-dessus. Les complexes ainsi formés sont révélés à l'aide d'un substrat approprié.

La présente invention a en outre pour objet un kit de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une paire d'amorces, une sonde, une puce à ADN ou à ARN, un vecteur recombinant, une cellule modifiée, une souche isolée de coronavirus, un polynucléotide, une protéine ou un peptide, un anticorps, et une puce à protéine tels que définis ci-dessus.

La présente invention a en outre pour objet, une composition immunogène, caractérisée en ce qu'elle comprend au moins un produit sélectionné dans le groupe constitué par :

- a) une protéine ou un peptide tels que définis ci-dessus,
- b) un polynucléotide de type ADN ou ARN ou l'un de ses fragments représentatifs tels que définis ci-dessus, de séquence choisie parmi :
  - (i) la séquence SEQ ID NO : 1 ou son équivalent ARN
  - (ii) la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,
  - (iii) la séquence complémentaire de la séquence SEQ ID NO : 1 ou de la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,
  - (iv) la séquence nucléotidique d'un fragment représentatif du polynucléotide tel que défini en (i), (ii) ou (iii),
  - (v) la séquence telle que définie en (i), (ii), (iii) ou (iv), modifiée, et

c) un vecteur d'expression recombinant comprenant un polynucléotide tel que défini en b), et

d) une banque d'ADNc telle que définie ci-dessus, ladite composition immunogène étant capable d'induire une immunité humorale ou  
 5 cellulaire protectrice spécifique du coronavirus associé au SRAS, notamment la production d'un anticorps dirigé contre un épitope spécifique du coronavirus associé au SRAS.

Les protéines et les peptides tels que définis ci-dessus, notamment les protéines S, M, E et/ou N et les peptides dérivés, ainsi que les molécules d'acide  
 10 nucléique (ADN ou ARN) codant lesdites protéines ou lesdits peptides, sont de bons candidats vaccin et peuvent être utilisées dans des compositions immunogènes pour la production d'un vaccin contre le coronavirus associé au SRAS.

Selon un mode de réalisation avantageux des compositions selon l'invention, elles contiennent en outre, au moins un véhicule pharmaceutiquement  
 15 acceptable et éventuellement des substances porteuses et/ou des adjuvants.

Les véhicules pharmaceutiquement acceptables, les substances porteuses et les adjuvants sont ceux classiquement utilisés.

Les adjuvants sont avantageusement choisis dans le groupe constitué par des émulsions huileuses, de la saponine, des substances minérales, des extraits  
 20 bactériens, de l'hydroxyde d'alumine et le squalène.

Les substances porteuses sont avantageusement sélectionnées dans le groupe constitué par des liposomes unilamellaires, des liposomes multilamellaires, des micelles de saponine ou des microsphères solides de nature saccharidique ou aurifère.

25 Les compositions selon l'invention, sont administrées par voie générale, notamment intramusculaire ou sous-cutanée ou bien par voie locale notamment nasale (aérosol).

La présente invention a également pour objet l'utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le  
 30 groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75 pour former un complexe immun avec un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet un complexe immunitaire formé d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet l'utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75 pour induire la production d'un anticorps capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet l'utilisation d'un polynuécléotide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 1, 2, 4, 7, 8, 13, 15, 16, 18, 19, 20, 31, 36 et 38 pour induire la production d'un anticorps dirigé contre la protéine codée par ledit polynuécléotide et capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS

Outre les dispositions qui précèdent, l'invention comprend encore d'autres dispositions, qui ressortiront de la description qui va suivre, qui se réfère à des exemples de mise en œuvre du polynuécléotide représentant le génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589, et des fragments d'ADNc dérivés objets de la présente invention, ainsi qu'au Tableau I présentant la liste des séquences :

**Tableau I : Liste des séquences**

numéro d'identification	Séquence	Position de l'ADNc en référence à Genbank AY274119.3	Numéro de dépôt à la CNCM du plasmide correspondant
SEQ ID NO : 1	génome de la souche issue du prélèvement 031589	-	-
SEQ ID NO : 2	ORF-S*	21406-25348	-
SEQ ID NO : 3	Protéine S	-	-
SEQ ID NO : 4	ORF-S**	21406-25348	I-3059
SEQ ID NO : 5	fragment Sa	21406-23454	I-3020

SEQ ID NO : 6	fragment Sb	23322-25348	I-3019
SEQ ID NO : 7	ORF-3+ORF-4*	25110-26244	-
SEQ ID NO : 8	ORF-3+ORF-4**	25110-26244	I-3126
SEQ ID NO : 9	ORF3	-	-
SEQ ID NO : 10	Protéine ORF-3	-	-
SEQ ID NO : 11	ORF4	-	-
SEQ ID NO : 12	Protéine ORF-4	-	-
SEQ ID NO : 13	ORF-E*	26082-26413	-
SEQ ID NO : 14	Protéine E	-	-
SEQ ID NO : 15	ORF-E**	26082-26413	I-3046
SEQ ID NO : 16	ORF-M*	26330-27098	-
SEQ ID NO : 17	Protéine M	-	-
SEQ ID NO : 18	ORF-M**	26330-27098	I-3047
SEQ ID NO : 19	ORF7 à 11*	26977-28218	-
SEQ ID NO : 20	ORF7 à 11**	26977-28218	I-3125
SEQ ID NO : 21	ORF7	-	-
SEQ ID NO : 22	Protéine ORF7	-	-
SEQ ID NO : 23	ORF8	-	-
SEQ ID NO : 24	Protéine ORF8	-	-
SEQ ID NO : 25	ORF9	-	-
SEQ ID NO : 26	Protéine ORF9	-	-
SEQ ID NO : 27	ORF10	-	-
SEQ ID NO : 28	Protéine ORF10	-	-
SEQ ID NO : 29	ORF11	-	-
SEQ ID NO : 30	Protéine ORF11	-	-
SEQ ID NO : 31	OrF1ab	265-21485	-
SEQ ID NO : 32	ORF13	28130-28426	-
SEQ ID NO : 33	Protéine ORF13	-	-
SEQ ID NO : 34	ORF14	-	-
SEQ ID NO : 35	Protéine ORF14	28583-28795	-
SEQ ID NO : 36	ORF-N*	28054-29430	-
SEQ ID NO : 37	Protéine N	-	-
SEQ ID NO : 38	ORF-N**	28054-29430	I-3048
SEQ ID NO : 39	5'non-codante**	1-204	I-3124
SEQ ID NO : 40	3'non-codante**	28933-29727	I-3123
	<b>ORF1ab</b>		-
SEQ ID NO : 41	Fragment L0	30-500	-
SEQ ID NO : 42	Fragment L1	211-2260	-
SEQ ID NO : 43	Fragment L2	2136-4187	-
SEQ ID NO : 44	Fragment L3	3892-5344	-
SEQ ID NO : 45	Fragment L4b	4932-6043	-
SEQ ID NO : 46	Fragment L4	5305-7318	-
SEQ ID NO : 47	Fragment L5	7275-9176	-
SEQ ID NO : 48	Fragment L6	9032-11086	-
SEQ ID NO : 49	Fragment L7	10298-12982	-
SEQ ID NO : 50	Fragment L8	12815-14854	-



SEQ ID NO : 51	Fragment L9	14745-16646	-
SEQ ID NO : 52	Fragment L10	16514-18590	-
SEQ ID NO : 53	Fragment L11	18500-20602	-
SEQ ID NO : 54	Fragment L12	20319-22224	-
SEQ ID NO : 55	Amorce N sens	-	-
SEQ ID NO : 56	Amorce N antisens	-	-
SEQ ID NO : 57	Amorce S <sub>C</sub> sens	-	-
SEQ ID NO : 58	Amorce S <sub>L</sub> sens	-	-
SEQ ID NO : 59	Amorce S <sub>C</sub> et S <sub>L</sub> antisens	-	-
SEQ ID NO : 60	Amorce sens série 1	28507-28522	-
SEQ ID NO : 61	Amorce antisens série 1	28774-28759	-
SEQ ID NO : 62	Amorce sens série 2	28375-28390	-
SEQ ID NO : 63	Amorce antisens série 2	28702-28687	-
SEQ ID NO : 64	Sonde 1/série 1	28561-28586	-
SEQ ID NO : 65	Sonde 2/série 1	28588-28608	-
SEQ ID NO : 66	Sonde 1/série 2	28541-28563	-
SEQ ID NO : 67	Sonde 2/série 2	28565-28589	-
SEQ ID NO : 68	Amorce ancre 14T		
SEQ ID NO : 69	Peptide M2-14	-	-
SEQ ID NO : 70	Peptide E1-12	-	-
SEQ ID NO : 71	Peptide E53-76	-	-
SEQ ID NO : 72	5'non-codante*	1-204	-
SEQ ID NO : 73	3'non-codante*	28933-29727	-
SEQ ID NO : 74	Protéine ORF1a	-	-
SEQ ID NO : 75	Protéine ORF1b	-	-
SEQ ID NO:76-139	Amorces		

\* produit d'amplification PCR (amplicon)

\*\* insert cloné dans le plasmide déposé à la CNCM

ainsi qu'aux dessins annexés dans lesquels :

- la figure 1 illustre l'analyse par Western-blot de l'expression *in vitro* des protéines recombinantes N, S<sub>C</sub> et S<sub>L</sub> à partir des vecteurs d'expression pIVEX. Piste 1 : pIV2.3N. Piste 2 : pIV2.3S<sub>C</sub>. Piste 3 : pIV2.3S<sub>L</sub>. Piste 4 : pIV2.4N. Piste 5 : pIV2.4S<sub>I</sub> ou pIV2.4S<sub>C</sub>. Piste 6 : pIV2.4S<sub>L</sub>. L'expression de la protéine GFP exprimée à partir du même vecteur est utilisée comme contrôle.

- la figure 2 illustre l'analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et coloration au bleu de

Coomassie, de l'expression *in vivo* de la protéine N à partir des vecteurs d'expression pIVEX. La souche d'*E.coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants est cultivée à 30°C dans du milieu LB, en présence ou en l'absence d'inducteur (IPTG 1mM). Piste 1 : pIV2.3N Piste 2 : pIV2.4N.

5 - la figure 3 illustre l'analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et coloration au bleu de Coomassie, de l'expression *in vivo* des polypeptides S<sub>L</sub> et S<sub>C</sub> à partir des vecteurs d'expression pIVEX. La souche d'*E.coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants est cultivée à 30°C dans du milieu LB, en présence ou  
10 en l'absence d'inducteur (IPTG 1mM). Piste 1 : pIV2.3S<sub>C</sub> Piste 2 : pIV2.3S<sub>L</sub>. Piste 3 : pIV2.4S<sub>I</sub> Piste 4 : pIV2.4S<sub>L</sub>.

- la figure 4 illustre l'activité antigénique des protéines N, S<sub>L</sub> et S<sub>C</sub> recombinantes produites dans la souche *E. coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants. A : électrophorèse (SDS-PAGE) des lysats bacté-  
15 riens. B et C : Western-blot avec les sérums, provenant d'un même patient infecté par le SARS-CoV, prélevés respectivement 8 jours (B : sérum M12) et 29 jours (C : sérum M13) après le début des symptômes du SRAS. Piste 1 : pIV2.3N. Piste 2 : pIV2.4N. Piste 3 : pIV2.3S<sub>C</sub>. Piste 4 : pIV2.4 S<sub>I</sub>. Piste 5 : pIV2.3S<sub>L</sub>. Piste 6 : pIV2.4S<sub>L</sub>

- la figure 5 illustre la purification sur colonne Ni-NTA agarose de la  
20 protéine N recombinante produite dans la souche *E. coli* BL21(DE3)pDIA17 à partir du vecteur pIV2.3N. Piste 1 : Extrait bactérien total. Piste 2 : Extrait soluble. Piste 3 : Extrait insoluble. Piste 4 : Extrait déposé sur la colonne Ni-NTA. Piste 5 : protéines non-retenues. Piste 6 : Fractions du pic 1. Piste 7 : Fractions du pic 2.

- la figure 6 illustre la purification de la protéine S<sub>C</sub> recombinante à  
25 partir des corps d'inclusions produits dans la souche *E. coli* BL21(DE3)pDIA17 transformée par le pIV2.4S<sub>I</sub>. A. Traitement au Triton X-100 (2%) : Piste 1 : Extrait bactérien total. Piste 2 : Extrait soluble. Piste 3 : Extrait insoluble. Piste 4 : Surnageant après traitement au Triton X-100 (2 %). Pistes 5 et 6 : Culot après traitement au Triton X-100 (2 %). B : Traitement à l'urée 4M, 5M, 6M et 7M des extraits solubles et inso-  
30 lubles.

- la figure 7 représente l'immunoempreinte réalisée à l'aide d'un lysat de cellules infectées par le SARS-CoV et d'un sérum de patient atteint de pneumopathie atypique.

5 - la figure 8 représente des immunoempreintes réalisées à l'aide d'un lysat de cellules infectées par le SARS-CoV et d'immunsérums de lapins spécifiques de la nucléoprotéine N (A) et de la protéine de spicule S (B). I.S. : sérum immun. p.i. : sérum pré-immun. L'immunsérum anti-N a été utilisé au 1/50000 et l'immunsérum anti-S au 1/10000.

10 - la figure 9 illustre la réactivité en ELISA des sérums polyclonaux monospécifiques de lapin dirigés contre la protéine N ou le fragment court de la protéine S (Sc), vis-à-vis des protéines recombinantes correspondantes utilisées pour l'immunisation. A : lapins P13097, P13081, et P13031 immunisés avec la protéine N recombinante purifiée. B : lapins P11135, P13042, et P14001 immunisés avec une  
15 (Sc). I.S. : sérum immun. p.i. : sérum pré-immun.

- la figure 10 illustre la réactivité en ELISA de la protéine N recombinante purifiée, vis-à-vis de sérum de patients atteints de pneumonie atypique causée par le SARS-CoV. Figure 10a : plaques ELISA préparés avec la protéine N à la concentration de 4 µg/ml et 2 µg/ml. Figure 10b : plaque ELISA préparée avec la  
20 protéine N à la concentration de 1 µg/ml. Les sérums désignés A, B, D, E, F, G, H correspondent à ceux du Tableau IV.

- la figure 11 illustre l'amplification par RT-PCR de quantités décroissantes d'ARN synthétique du gène N du SARS-CoV ( $10^7$  à 1 copie), à l'aide des couples d'amorces n° 1 (N/+28507,N/-28774) (A) et n° 2 (N/+28375,N/-28702)  
25 (B). T : amplification réalisée en l'absence d'ARN. MW : marqueur d'ADN.

- la figure 12 illustre l'amplification par RT-PCR en temps réel d'ARN synthétique du gène N du SARS-CoV : des quantités décroissantes d'ARN synthétique en répliquat (repli. ; pistes 16 à 29) ainsi que de l'ARN viral dilué au  $1/20 \times 10^{-4}$  (piste 32) ont été amplifiés par RT-PCR en temps réel à l'aide du kit "Light  
30 Cycler RNA Amplification Kit Hybridization Probes" et des couples d'amorces et de sondes de la série n° 2, dans les conditions décrites à l'exemple 7.

- la figure13 (figure 13.1 à 13.70) représente la carte de restriction de la séquence SEQ ID NO : 1 correspondant à l'équivalent ADN du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589.

Il doit être bien entendu, toutefois, que ces exemples sont donnés  
5 uniquement à titre d'illustration de l'objet de l'invention, dont ils ne constituent en aucune manière une limitation.

**Exemple 1 : Clonage et séquençage du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589**

L'ARN de la souche de SARS-CoV a été extrait à partir du prélè-  
10 vement de lavage bronchoalvéolaire répertorié sous le numéro 031589, effectué sur un patient de l'hôpital français de Hanoi (Vietnam) atteint de SRAS.

L'ARN isolé a été utilisé comme matrice pour amplifier les ADNc correspondant aux différents cadres ouverts de lecture du génome (ORF 1a, ORF1b, ORF-S, ORF-E, ORF-M, ORF-N (incluant les ORF-13 et ORF-14), ORF3, ORF4,  
15 ORF7 à ORF11), et aux extrémités 5' et 3' non-codantes. Les séquences des amorces et des sondes utilisées pour l'amplification/détection ont été définies d'après la séquence nucléotidique disponible du SARS-CoV.

Dans ce qui suit les amorces et les sondes sont identifiées par : la lettre S, suivie d'une lettre qui indique la région correspondante du génome (L pour  
20 l'extrémité 5' incluant ORF1a et ORF1b ; S, M et N pour les ORF-S, ORF-M, ORF-N, SE et MN pour les régions intergéniques correspondantes), puis éventuellement de Fn, Rn, avec n inclus entre 1 et 6 correspondant aux amorces utilisées pour la PCR nichée ou imbriquée (paire F1 + R1 pour la première amplification, paire F2 + R2 pour la deuxième amplification, etc...), puis de +/- ou -/- correspondant à une amorce sens ou  
25 antisens et enfin des positions des amorces en référence à la séquence Genbank AY27411.3 ; pour les amorces S et N sens et antisens et les autres amorces sens uniquement, lorsqu'une seule position est indiquée elle correspond à celle de l'extrémité 5' d'une sonde ou d'une amorce d'environ 20 bases ; pour les amorces antisens autres que les amorces S et N, lorsqu'une seule position est indiquée elle  
30 correspond à celle de l'extrémité 3' d'une sonde ou d'une amorce d'environ 20 bases.

Les produits d'amplifications ainsi générés ont été séquencés à l'aide d'amorces spécifiques afin de déterminer la séquence complète du génome de la

souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589. Ces produits d'amplification, à l'exception de ceux correspondant aux ORF1a et ORF1b, ont ensuite été clonés dans des vecteurs d'expression afin de produire les protéines virales correspondantes et les anticorps dirigés contre ces protéines, notamment par immunisation à base d'ADN.

### 1. Extraction des ARN

Les ARN ont été extraits à l'aide du kit *QIamp viral RNA extraction mini* (QIAGEN) en suivant les recommandations du fabricant. De manière plus précise : 140 µl du prélèvement et 560 µl de tampon AVL ont été mélangés vigoureusement pendant 15 secondes, incubés 10 min à température ambiante puis centrifugés brièvement à vitesse maximale. 560 µl d'éthanol à 100% ont été ajoutés au surnageant et le mélange ainsi obtenu a été agité très vigoureusement pendant 15 sec. 630 µl du mélange ont ensuite été déposés sur la colonne.

La colonne a été placée sur un tube de 2 ml, centrifugée 1 min à 8000 rpm, puis le reste du mélange précédent a été déposé sur la même colonne, centrifugé à nouveau, 1 min à 8000 rpm et la colonne a été transférée sur un tube de 2 ml propre. Ensuite, 500 µl de tampon AW1 ont été ajoutés sur la colonne, puis la colonne a été centrifugée 1 min à 8000 rpm et l'éluat a été éliminé. 500 µl de tampon AW2 ont été ajoutés sur la colonne qui a ensuite été centrifugée 3 min à 14000 rpm et transférée sur un tube de 1,5 ml. Enfin, 60 µl de tampon AVE ont été ajoutés sur la colonne qui a été incubée 1 à 2 min à température ambiante puis centrifugée 1 min à 8000 rpm. L'éluat correspondant à l'ARN purifié a été récupéré et congelé à -20°C.

### 2. Amplification, séquençage et clonage des ADNc

#### 2.1) ADNc codant pour la protéine S

Les ARN extraits à partir du prélèvement ont été soumis à une transcription inverse à l'aide d'oligonucléotides hexamériques de séquence aléatoire (pdN6), afin de produire des fragments d'ADNc.

La séquence codant pour la glycoprotéine S du SARS-CoV a été amplifiée sous la forme de deux fragments d'ADN chevauchants : fragment 5' (SRAS-Sa, SEQ ID NO:5) et fragment 3' (SRAS-Sb, SEQ ID NO:6), en réalisant deux amplifications successives à l'aide d'amorces imbriquées. Les amplicons ainsi obtenus

ont été séquencés, clonés dans le vecteur plasmidique PCR 2.1-TOPO™ (IN VITROGEN), puis la séquence des ADNc clonés a été déterminée.

a) clonage et séquençage des fragments Sa et Sb

a<sub>1</sub>) synthèse de l'ADNc

5 Le mélange réactionnel contenant : ARN (5 µl) , H<sub>2</sub>O ppi (3,5 µl), tampon de transcriptase inverse 5X (4 µl), dNTP 5 mM (2 µl), pdN6 100 ug/ml (4 µl), RNasin 40 UI/ul (0,5 µl) et transcriptase inverse AMV-RT, 10 UI/ul, PROMEGA (1µl) a été incubé dans un thermocycleur dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis l'ADNc obtenu a été maintenu à +4°C.

10 a<sub>2</sub>) première amplification PCR

Les extrémités 5' et 3' du gène S ont été amplifiées respectivement avec les paires d'amorces S/F1/+ 21350-21372 et S/R1/- 23518-23498, S/F3/+ 23258-23277 et S/R3/- 25382-25363. Le mélange réactionnel de 50 µl contenant : ADNc (2 µl), amorces 50 µM (0,5 µl), tampon 10 X (5 µl), dNTP 5 mM (2 µl), Taq Expand High Fidelity, Roche (0,75 µl) et H<sub>2</sub>O (39, 75 µl) a été amplifié dans un thermocycleur, dans les conditions suivantes : une étape initiale de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant : une étape de dénaturation à 94°C pendant 30 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 2 min 30 sec, avec 10 sec d'élongation supplémentaire à chaque cycle, puis d'une étape finale d'élongation à 72°C pendant 5 min.

20 a<sub>3</sub>) deuxième amplification PCR

Les produits de la première amplification PCR (amplicons 5' et 3') ont subi une seconde étape d'amplification PCR (PCR nichée) dans des conditions identiques à celles de la première amplification, avec les paires d'amorces S/F2/+ 21406-21426 et S/R2/- 23454-23435, et S/F4/+ 23322-23341 et S/R4/- 25348-25329, respectivement pour l'amplicon 5' et l'amplicon 3'.

a<sub>4</sub>) clonage et séquençage des fragments Sa et Sb

30 Les amplicons Sa (extrémité 5') et Sb (extrémité 3') ainsi obtenus ont été purifiés à l'aide du kit *QIAquick PCR purification* (QIAGEN), en suivant les recommandations du fabricant, puis ils ont été clonés dans le vecteur PCR2.1-TOPO (kit Invitrogen), pour donner les plasmides dénommés SRAS-S1 et SRAS-S2.

L'ADN des clones Sa et Sb a été isolé puis l'insert correspondant a été séquencé à l'aide du Kit Big Dye, Applied Biosystem® et des amorces universelles M13 forward et M13 reverse, ainsi que des amorces : S/S+/21867, S/S+/22353, S/S+/22811, S/S+/23754, S/S+/24207, S/S+/24699, S/S+/24348, S/S-/24209, S/S-/23630, S/S-/23038, S/S-/22454, S/S-/21815, S/S-/24784, S/S+/21556, S/S+/23130 et S/S+/24465, en suivant les instructions du fabricant ; les séquences des fragments Sa et Sb ainsi obtenues correspondent aux séquences SEQ ID NO :5 et SEQ ID NO :6 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-S1 a été déposé sous le n° I-3020, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 5' de la séquence du gène S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment dénommé Sa correspondant aux nucléotides des positions 21406 à 23454 (SEQ ID NO :5), en référence à la séquence Genbank AY274119.3 Tor2.

Le plasmide, dénommé TOP10F'-SARS-S2 a été déposé sous le n° I-3019, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 3' de la séquence du gène S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment dénommé Sb correspondant aux nucléotides des positions 23322 à 25348 (SEQ ID NO : 6), en référence à la séquence Genbank n° d'accès AY274119.3.

**b) clonage et séquençage de l'ADNc complet (clone SRAS-S de 4 kb)**

L'ADNc S complet a été obtenu à partir des clones SARS-S1 et SARS-S2 précités, de la façon suivante :

- 1) une réaction d'amplification PCR a été réalisée sur un clone SARS-S2 en présence de l'amorce S/R4-/25348-25329 précitée et de l'amorce S/S+/24696-24715: un amplicon de 633 bp a été obtenu,
- 2) une autre réaction d'amplification PCR a été réalisée sur un autre clone SARS-S2, en présence des amorces S/F4+/23322-23341 précitée et S/S-/24803-24784: un amplicon de 1481 pb a été obtenu,

La réaction d'amplification a été réalisée dans les conditions telles que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception que 30 cycles d'amplifications comprenant une étape de dénaturation à 94° C pendant 20 sec et une étape d'élongation à 72° C pendant 2 min 30 sec ont été effectués.

5 3) les 2 amplicons (633 pb et 1481 pb) ont été purifiés dans les conditions telles que définies ci-dessus pour les fragments Sa et Sb.

4) une autre réaction d'amplification PCR à l'aide des amorces S/F4+/23322-23341 et S/R4-/25348-25329 précitées, a été réalisée sur les amplicons purifiés obtenus en 3). La réaction d'amplification a été réalisée dans les conditions  
10 telles que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception que 30 cycles d'amplifications ont été effectués.

L'amplicon de 2026 pb ainsi obtenu a été purifié, cloné dans le vecteur PCR2.1-TOPO puis séquencé comme ci-dessus, à l'aide des amorces telles que définies ci-dessus pour les fragments Sa et Sb. Le clone ainsi obtenu a été  
15 dénommé clone 3'.

5) Le clone SARS-S1 précédemment obtenu et le clone 3' ont été digérés par *EcoR I*, les bandes d'environ 2kb ainsi obtenues ont été purifiées sur gel puis amplifiées par PCR avec les amorces S/F2+/21406-21426 et S/R4-/25348-25329 précitées. La réaction d'amplification a été réalisée dans les conditions telles  
20 que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception que 30 cycles d'amplifications ont été effectués. L'amplicon d'environ 4 kb a été purifié et séquencé. Il a ensuite été cloné dans le vecteur PCR2.1-TOPO pour donner le plasmide, dénommé SARS-S, et l'insert contenu dans ce plasmide a été séquencé comme ci-dessus, à l'aide des amorces telles que définies ci-dessus pour les fragments  
25 Sa et Sb. Les séquences d'ADNc de l'insert et de l'amplicon codant pour la protéine S, correspondent respectivement aux séquences SEQ ID NO : 4 et SEQ ID NO : 2 dans la liste de séquences jointe en annexe, elles codent pour la protéine S (SEQ ID NO : 3).

La séquence de l'amplicon correspondant à l'ADNc codant pour la  
30 protéine S de la souche de SARS-CoV issue du prélèvement n°031589 présente les deux mutations suivantes par rapport aux séquences correspondantes de respective-



ment les isolats Tor2 et Urbani, les positions des mutations étant indiquées en référence à la séquence complète du génome de l'isolat Tor2 (Genbank AY274119.3) :

- g/t en position 23220 ; le codon alanine (gct) en position 577 de la séquence en acides aminés de la protéine S de Tor2 est remplacé par un codon sérine (tct),

- c/t en position 24872 : cette mutation ne modifie pas la séquence en acides aminés de la protéine S, et

Le plasmide, dénommé SARS-S, a été déposé sous le n° I-3059, le 20 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, laquelle séquence correspondant aux nucléotides des positions 21406 à 25348 (SEQ ID NO :4), en référence à la séquence Genbank AY274119.3.

## 2.2) ADNc codant pour les protéines M et E

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse, associée, lors de la même étape (kit *Titan One Step RT-PCR*®, Roche), à une réaction d'amplification par PCR, à l'aide des couples d'amorces :

- S/E/F1/+ /26051-26070 et S/E/R1/- /26455-26436 pour amplifier l'ORF-E, et
- S/M/F1/+ /26225-26244 et S/M/R1/- /27148-27129 pour amplifier l'ORF-M.

Un premier mélange réactionnel contenant : 8,6 µl d'H<sub>2</sub>Oppi, 1 µl de dNTP (5mM), 0,2 µl de chacune des amorces (50µM), 1,25 µl de DTT (100mM) et 0,25 µl de RNAsin (40UI/µl) a été combiné avec un deuxième mélange réactionnel contenant : 1 µl d'ARN, 7 µl d'H<sub>2</sub>Oppi, 5 µl de tampon de RT-PCR 5X et 0,5 µl de mélange d'enzyme et les mélanges combinés ont été incubés dans un thermocycleur dans les conditions suivantes : 30 min à 42°C, 10 min à 55°C, 2 min à 94°C suivi de 40 cycles comprenant une étape de dénaturation à 94°C pendant 10 sec, une étape d'hybridation à 55°C pendant 30 sec et une étape d'élongation à 68°C pendant 45 sec, avec 3 sec d'incrément par cycle et enfin une étape d'élongation terminale à 68°C pendant 7 min.

Les produits d'amplification ainsi obtenus (amplicons M et E) ont subi une deuxième amplification PCR (PCR nichée) en utilisant le kit Expand High-Fi®, Roche), à l'aide des couples d'amorces :

- S/E/F2/+ /26082-26101 et S/E/R2/- /26413-26394 pour l'amplicon E, et
- 5 - S/M/F2/+ /26330-26350 et S/M/R2/- /27098-27078 pour l'amplicon M.

Le mélange réactionnel contenant : 2 µl du produit de la première PCR, 39,25 µl d'H<sub>2</sub>O<sub>pp</sub>i, 5 µl de tampon 10X contenant du MgCl<sub>2</sub>, 2 µl de dNTP (5mM), 0,5µl de chacune des amorces (50 µM) et 0,75µl de mélange d'enzyme a été incubé dans un thermocycleur dans les conditions suivantes : une étape de dénaturation à 94°C pendant 2 min a été suivie de 30 cycles comprenant une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 60°C pendant 30 sec et une étape d'élongation à 72°C pendant 45 sec, avec 3 sec d'incrément par cycle, et enfin une étape d'élongation terminale à 72°C pendant 7 min. Les produits d'amplification obtenus correspondant aux ADNc codant pour les protéines E et M ont été séquencés

10 comme ci-dessus, à l'aide des amorces : S/E/F2/+ /26082 et S/E/R2/- /26394, S/M/F2/+ /26330, S/M/R2/- /27078 précitées et des amorces S/M/+ /26636-26655 et S/M/- /26567-26548. Ils ont ensuite été clonés, comme ci-dessus, pour donner les plasmides dénommés SARS-E et SARS-M. L'ADN de ces clones a ensuite été isolé et séquencé à l'aide des amorces universelles M13 forward et M13 reverse ainsi que des

15 amorces S/M/+ /26636 et S/M/- /26548 précitées.

20

La séquence de l'amplicon représentant l'ADNc codant pour la protéine E (SEQ ID NO : 13) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani. La séquence de la protéine E de la souche de SARS-CoV 031589 correspond à la séquence SEQ ID NO : 14 dans la

25 liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-E a été déposé sous le n° I-3046, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant

30 pour la protéine E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux

nucléotides des positions 26082 à 26413 (SEQ ID NO :15), en référence à la séquence Genbank n° d'accès AY274119.3.

La séquence de l'amplicon représentant l'ADNc codant pour la M (SEQ ID NO :16) de la souche de SARS-CoV issue du prélèvement n°031589 ne  
 5 comporte pas de différences par rapport à la séquence correspondante de l'isolat AY274119.3-Tor2. En revanche, en position 26857, l'isolat AY278741-Urbani comporte un c et la séquence de la souche de SARS-CoV issue du prélèvement répertorié sous le n°031589 un t. Cette mutation aboutit à une modification de la séquence en acides aminés de la protéine correspondante: en position 154, une proline  
 10 (AY278741-Urbani) est changée en sérine dans la souche de SARS-CoV issue du prélèvement répertorié sous le n°031589. La séquence de la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n°031589 correspond à la séquence SEQ ID NO :17 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-M a été déposé sous le n° I-3047, le  
 15 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus ; laquelle séquence correspondant aux nucléotides des positions 26330 à 27098 (SEQ ID NO :18), en référence à la séquence  
 20 Genbank n° d'accès AY274119.3.

### **2.3) ADNc correspondant aux ORF3, ORF4, ORF7 à ORF11**

La même stratégie d'amplification, de clonage et de séquençage a été utilisée pour obtenir les fragments d'ADNc correspondant respectivement aux ORF suivantes : ORF 3, ORF4, ORF7, ORF8, ORF9, ORF10 et ORF11. Les couples  
 25 d'amorces utilisées pour la première amplification sont :

- ORF3 et ORF4 : S/SE/F1/+ /25069-25088 et S/SE/R1/- /26300-26281
- ORF7 à ORF11 : S/MN/F1/+ /26898-26917 et S/MN/R1/- /28287-28266

Les couples d'amorces utilisées pour la deuxième amplification sont :

- ORF3 et ORF4 : S/SE/F2/+ /25110-25129 et S/SE/R2/- /26244-26225
- 30 - ORF7 à ORF11 : S/MN/F2/+ /26977-26996 et S/MN/R2/- /28218-28199

Les conditions de la première amplification (RT-PCR) sont les suivantes : 45 min à 42°C, 10 min à 55°C, 2 min à 94°C suivi de 40 cycles compre-

nant une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 58°C pendant 30 sec et une étape d'élongation à 68°C pendant 1 min, avec 5 sec d'incrément par cycle et enfin une étape d'élongation terminale à 68°C pendant 7 min.

Les conditions de la PCR nichée sont les suivantes : une étape de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 58°C pendant 30 sec et une étape d'élongation à 72°C pendant 50 sec, avec 4 sec d'incrément par cycle et enfin une étape d'élongation terminale à 72°C pendant 7 min.

Les produits d'amplification obtenus correspondant aux ADNc contenant respectivement les ORF3 et 4 et les ORF7 à 11 ont été séquencés à l'aide des amorces : S/SE+/25363, S/SE+/25835, S/SE-/25494, S/SE-/25875, S/MN+/27839, S/MN+/27409, S/MN-/27836 S/MN-/27799 et clonés comme ci-dessus pour les autres ORF, pour donner les plasmides dénommés SARS-SE et SARS-MN. L'ADN de ces clones a été isolé et séquencé à l'aide de ces mêmes amorces et des amorces universelles M13 sens et M13 anti-sens.

La séquence de l'amplicon représentant l'ADNc de la région contenant les ORF 3 et 4 (SEQ ID NO :7) de la souche de SARS-CoV issue du prélèvement n°031589 comporte une différence nucléotidique par rapport à la séquence correspondante de l'isolat AY274119-Tor2. Cette mutation en position 25298 aboutit à une modification de la séquence en acides aminés de la protéine correspondante (ORF 3): en position 11, une arginine (AY274119-Tor2) est changée en glycine dans la souche de SARS-CoV issue du prélèvement n°031589. En revanche, aucune mutation n'a été identifiée par rapport à la séquence correspondante de l'isolat AY278741-Urbani. Les séquences des ORF 3 et 4 la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO :10 et 12 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-SE a été déposé sous le n° I-3126, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc correspondant à la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle région correspondant aux nucléotides des

positions 25110 à 26244 (SEQ ID NO :8), en référence à la séquence Genbank n° d'accès AY274119.3,

La séquence de l'amplicon représentant l'ADNc correspondant à la région contenant les ORF7 à ORF11 (SEQ ID NO :19) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119-Tor2 et AY278741-Urbani. Les séquences des ORF7 à 11 de la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO : 22, 24, 26, 28 et 30 dans la liste de séquences jointe en annexe.

Le plasmide dénommé SARS-MN a été déposé sous le n° I-3125, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc correspondant à la région située entre l'ORF-M et l'ORF-N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 et prélevée à Hanoi, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 26977 à 28218 (SEQ ID NO :20 ), en référence à la séquence Genbank n° d'accès AY274119.3.

La séquence de l'amplicon représentant l'ADNc correspondant à la région contenant les ORF7 à ORF11 (SEQ ID NO :19) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119-Tor2 et AY278741-Urbani. Les séquences des ORF7 à 11 de la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO : 22, 24, 26, 28 et 30 dans la liste de séquences jointe en annexe.

#### 2.4) ADNc codant pour la protéine N et incluant les ORF13 et ORF14

L'ADNc a été synthétisé et amplifié comme décrit ci-dessus pour les fragments Sa et Sb. De manière plus précise, le mélange réactionnel contenant : 5 µl d'ARN, 5 µl d'H<sub>2</sub>O ppi 4 µl de tampon de reverse transcriptase 5X, 2 µl de dNTP (5 mM), 2 µl d'oligo 20T (5 µM), 0,5 µl de RNasin (40 UI/ul) et 1, 5 µl de AMV-RT (10 UI/ul Promega) a été incubé dans un thermocycleur dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis il a été maintenu à +4°C.

Une première amplification PCR a été réalisée avec la paire d'amorces S/N/F3/+28023 et S/N/R3/-29480.

Le mélange réactionnel comme ci-dessus pour l'amplification des fragments S1 et S2 a été incubé dans un thermocycleur, dans les conditions suivantes :

- 5 une étape initiale de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 1 min 30 sec avec 10 sec d'élongation supplémentaire à chaque cycle, puis d'une étape finale d'élongation à 72°C pendant 5 min.

- 10 L'amplicon obtenu à la première amplification PCR a subi une seconde étape d'amplification PCR (PCR nichée) avec la paires d'amorce S/N/F4/+28054 et S/N/R4/-29430 dans des conditions identiques à celles de la première amplification.

- 15 Le produit d'amplification obtenu correspondant à l'ADNc codant pour la protéine N de la souche de SARS-CoV issue du prélèvement n°031589a été séquencé à l'aide des amorces : S/N/F4/+28054, S/N/R4/-29430, S/N/+28468, S/N/+28918 et S/N/-28607 et cloné comme ci-dessus pour les autres ORF, pour donner le plasmide dénommé SARS-N. L'ADN de ces clones a été isolé et séquencé à l'aide des amorces universelles M13 sens et M13 anti-sens, ainsi que des amorces
- 20 S/N/+28468, S/N/+28918 et S/N/-28607.

- La séquence de l'amplicon représentant l'ADNc correspondant à l'ORF-N et incluant les ORF13 et ORF14 (SEQ ID NO :36) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani. La
- 25 séquence de la protéine N de la souche de SARS-CoV issue du prélèvement n°031589 correspond à la séquence SEQ ID NO : 37 dans la liste de séquences jointe en annexe.

Les séquences des ORF13 et 14 de la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO : 32 et 34 dans la liste de séquences jointe en annexe.

- 30 Le plasmide dénommé SARS-N a été déposé sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc codant pour la protéine

N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 28054 à 29430 (SEQ ID NO :38), en référence à la séquence Genbank n° d'accès AY274119.3.

## 5 2.5) extrémités 5' et 3' non-codantes

### a) extrémité 5' non-codante (5'NC)

#### a<sub>1</sub>) synthèse de l'ADNc

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse dans les conditions suivantes :

10 L'ARN (15 µl) et l'amorce S/L/-/443 (3 µl à la concentration de 5µm, ont été incubés 10 min à 75°C.

Ensuite, du Tampon de transcriptase inverse 5X (6 µl, INVITROGEN), des dNTP 10 mM (1 µl), du DTT 0,1M (3 µl) ont été ajoutés et le mélange a été incubé à 50°C pendant 3 min.

15 Enfin la transcriptase inverse (3 µl de Superscript®, INVITROGEN) a été ajoutée au mélange précédent qui a été incubé à 50°C pendant 1h30 puis à 90 °C pendant 2 min.

L'ADNc ainsi obtenu a été purifié à l'aide du kit *QIAquick PCR purification* (QIAGEN), selon les recommandations du fabricant.

#### 20 b<sub>1</sub>) Réaction à la Terminal Transferase (TdT)

L'ADNc (10 µl) est incubé 2 min à 100°C, conservé dans la glace, puis sont ajoutés : H<sub>2</sub>O (2,5 µl), tampon TdT 5X (4 µl, AMERSHAM), dATP 5mM (2 µl) et TdT (1,5 µl, AMERSHAM). Le mélange ainsi obtenu est incubé 45 min à 37°C puis 2 min à 65°C.

25 Le produit obtenu est amplifié par une première réaction PCR à l'aide des amorces : S/L/-/225-206 et ancre 14T : 5'-AGATGAATTCGGTACCTTTTTTTTTTTTTTTT-3' (SEQ ID NO :68). Les conditions de l'amplification sont les suivantes : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 10 cycles comprenant une étape de dénaturation à 94°C pendant 10 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec puis de 30 cycles comprenant une étape de dénaturation à 94°C pendant 10 sec, une étape d'hybridation à 50°C pendant 30 sec puis

30

une étape d'élongation à 72°C pendant 30 sec, puis d'une étape finale d'élongation à 72°C pendant 5 min.

Le produit de la première amplification PCR a subi une seconde étape d'amplification à l'aide des amorces : S/L/-/204-185 et ancre 14T précitée dans des conditions identiques à celles de la première amplification. L'amplicon ainsi obtenu a été purifié, séquencé à l'aide de l'amorce S/L/-/182-163 puis il a été cloné comme ci-dessus pour les différentes ORF, pour donner le plasmide dénommé SARS-5'NC. L'ADN de ce clone a été isolé et séquencé à l'aide des amorces universelles M13 sens et M13 anti-sens et de l'amorce S/L/-/182-163 précitée.

L'amplicon représentant l'ADNc correspondant à l'extrémité 5'NC de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 correspond à la séquence SEQ ID NO : 72 dans la liste de séquences jointe en annexe ; cette séquence ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani.

Le plasmide dénommé SARS-5'NC a été déposé sous le n° I- 3124, le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc correspondant à l'extrémité 5'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 1 à 204 (SEQ ID NO :39 ), en référence à la séquence Genbank n° d'accès AY274119.3.

**b) extrémité 3'non-codante (3'NC)**

**a<sub>1</sub>) synthèse de l'ADNc**

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse, selon le protocole suivant : le mélange réactionnel contenant : ARN (5 µl), H<sub>2</sub>O (5 µl), tampon de transcriptase inverse 5X (4 µl), dNTP 5 mM (2 µl), Oligo 20T 5µM (2 µl), RNasin 40 U/ µl (0,5 µl) et RT-AMV 10 UI/ µl (1,5 µl, PROMEGA) a été incubé dans un thermocycleur, dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis il a été maintenu à +4°C.

L'ADNc obtenu a été amplifié par une première réaction PCR à l'aide des amorces S/N+/28468-28487 et ancre 14T précitée. Les conditions de



l'amplification sont les suivantes : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 10 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 50 sec puis de 30 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 50°C pendant 30 sec puis une étape d'élongation à 72°C pendant 50 sec, puis d'une étape finale d'élongation à 72°C pendant 5 min.

Le produit de la première amplification PCR a subi une seconde étape d'amplification à l'aide des amorces S/N/+28933-28952 et ancre 14T précitée, dans des conditions identiques à celles de la première amplification. L'amplicon ainsi obtenu a été purifié, séquencé à l'aide de l'amorce S/N/+29257-29278 et cloné comme ci-dessus pour les différentes ORF, pour donner le plasmide dénommé SARS-3'NC. L'ADN de ce clone a été isolé et séquencé à l'aide des amorces universelles M13 sens et M13 anti-sens et de l'amorce S/N/+29257-29278 précitée.

L'amplicon représentant l'ADNc correspondant à l'extrémité 3'NC de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 correspond à la séquence SEQ ID NO :73 dans la liste de séquences jointe en annexe ; cette séquence ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani.

Le plasmide dénommé SARS-3'NC a été déposé sous le n° I-3123 le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15. ; il contient la séquence d'ADNc correspondant à l'extrémité 3'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant à celle située entre le nucléotide en position 28933 à 29727 (SEQ ID NO :40), en référence à la séquence Genbank n° d'accès AY274119.3, se termine par une série de nucléotides a.

## 2.6) ORF1a et ORF1b

L'amplification de la région 5' contenant les ORF1a et ORF1b du génome du SARS-CoV issu du prélèvement 031589 a été réalisée en pratiquant des réactions de RT-PCR suivies de PCR nichées selon les mêmes principes que ceux précédemment décrits pour les autres ORF. Les fragments amplifiés sont chevaux-

chants sur plusieurs dizaines de bases, permettant ainsi la reconstruction informatique de la séquence complète de cette partie du génome. En moyenne, les fragments amplifiés sont de deux kilobases.

- 14 fragments chevauchants dénommés L0 à L12 ont ainsi été
- 5 amplifiés à l'aide des amorces suivantes :

**Tableau II: Amorces utilisées pour l'amplification de la région 5'(ORF1a et ORF1b)**

REGION AMPLIFIEE ET SEQUENCEE (ne tient pas compte des amorces)	Amorce sens RT-PCR	Amorce antisens RT-PCR	Amorce sens PCR nichée	Amorce antisens PCR nichée
L0 50-480	S/L0/F1/+30	S/L0/R1/-481		
L1 231-2240	S/L1/F1/+147	S/L1/R1/-2336	S/L1/F2/+211	S/L1/R2/-2241
L2 2156-4167	S/L2/F1/+2033	S/L2/R1/-4192	S/L2/F2/+2136	S/L2/R2/-4168
L3 3913-5324	S/L3bis/F1/+3850	S/L3bis/R1/-5365	S/L3bis/F2/+3892	S/L3bis/R2/-5325
L4b 4952-6023	S/L4b/F1/+4878	S/L4b/R1/-6061	S/L4b/F2/+4932	S/L4b/R2/-6024
L4 5325-7318	S/L4/F1/+5272	S/L4/R1/-7392	S/L4/F2/+5305	S/L4/R2/-7323
L5 7296-9156	S/L5/F1/+7111	S/L5/R1/-9253	S/L5/F2/+7275	S/L5/R2/-9157
L6 9053-11066	S/L6/F1/+8975	S/L6/R1/-11151	S/L6/F2/+9032	S/L6/R2/-11067
L7 10928-12962	S/L7/F1/+10883	S/L7/R1/-13050	S/L7/F2/+10928	S/L7/R2/-12963
L8 12835-14834	S/L8/F1/+12690	S/L8/R1/-14857	S/L8/F2/+12815	S/L8/R2/-14835
L9 14765-16624	S/L9/F1/+14688	S/L9/R1/-16678	S/L9/F2/+14745	S/L9/R2/-16625
L10 16534-18570	S/L10/F1/+16451	S/L10/R1/-18594	S/L10/F2/+16514	S/L10/R2/-18571
L11 18521-20582	S/L11/F1/+18441	S/L11/R1/-20612	S/L11/F2/+18500	S/L11/R2/-20583
L12 20338-22205	S/L12/F1/+20279	S/L12/R1/-22229	S/L12/F2/+20319	S/L12/R2/-22206

Tous les fragments ont été amplifiés dans les conditions suivantes, excepté le fragment L0 qui a été amplifié comme décrit ci-dessus pour l'ORF-M :

- 10 - RT-PCR: 30 min à 42°C, 15 min à 55°C, 2 min à 94°C, puis l'ADNc obtenu est amplifié dans les conditions suivantes : 40 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 58°C pendant 30 sec puis une étape d'élongation à 68°C pendant 1 min 30 sec, avec 5 sec d'élongation supplémentaire à chaque cycle, puis une étape finale d'élongation à 68°C
- 15 pendant 7 min.

- PCR nichée : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 35 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 60°C pendant 30 sec puis une étape d'élongation à 72°C pendant 1 min 30 sec, avec 5 sec d'élongation supplémentaire à chaque cycle,
- 5 puis une étape finale d'élongation à 72°C pendant 7 min.

Les produits d'amplifications ont été séquencés à l'aide des amorces définies dans le Tableau III ci-après :

**Tableau III : Amorces utilisées pour le séquençage de la région 5' (ORF1a et ORF1b)**

10

Noms	Séquences (SEQ ID NO : 76 à 139)
S/L3/+14932	5'-CCACACACAGCTTGTGGATA-3'
S/L4/+16401	5'-CCGAAGTTGTAGGCAATGTC-3'
S/L4/+16964	5'-TTTGGTGCTCCTTCTTATTG-3'
S/L4/-16817	5'-CCGGCATCCAAACATAATTT-3'
S/L5/-17633	5'-TGGTCAGTAGGGTTGATTGG-3'
S/L5/-18127	5'-CATCCTTTGTGTCAACATCG-3'
S/L5/-18633	5'-GTCACGAGTGACACCATCCT-3'
S/L5/+17839	5'-ATGCGACGAGTCTGCTTCTA-3'
S/L5/+18785	5'-TTCATAGTGCCTGGCTTACC-3'
S/L5/+18255	5'-ATCTTGGCGCATGTATTGAC-3'
S/L6/-19422	5'-TGCATTAGCAGCAACAACAT-3'
S/L6/-19966	5'-TCTGCAGAACAGCAGAAGTG-3'
S/L6/-110542	5'-CCTGTGCAGTTTGTCTGTCA-3'
S/L6/+110677	5'-CCTTGTGGCAATGAAGTACA-3'
S/L6/+110106	5'-ATGTCATTTGCACAGCAGAA-3'
S/L6/+19571	5'-CTTCAATGGTTTGCCATGTT-3'
S/L7/-111271	5'-TGCGAGCTGTCATGAGAATA-3'
S/L7/-111801	5'-AACCGAGAGCAGTACACACAG-3'
S/L7/-112383	5'-TTTGGCTGCTGTAGTCAATG-3'
S/L7/+112640	5'-CTACGACAGATGTCCTGTGC-3'
S/L7/+112088	5'-GAGCAGGCTGTAGCTAATGG-3'
S/L7/+111551	5'-TTAGGCTATTGTTGCTGCTG-3'
S/L8/-13160	5'-CAGACAACATGAAGCACCAC-3'
S/L8/-113704	5'-CGCTGACGTGATATATGTGG-3'
S/L8/-14284	5'-TGCACAATGAAGGATACACC-3'
S/L8/+114453	5'-ACATAGCTCGCGTCTCAGTT-3'
S/L8/+113968	5'-GGCATTGTAGGCGTACTGAC-3'
S/L8/+113401	5'-GTTTGCGGTGTAAGTGCAG-3'
S/L9/-15098	5'-TAGTGGCGGCTATTGACTTC-3'
S/L9/-15677	5'-CTAAACCTTGAGCCGCATAG-3'
S/L9/-16247	5'-CATGGTCATAGCAGCACTTG-3'
S/L9/+16323	5'-CCAGGTTGTGATGTCACTGAT-3'
S/L9/+15858	5'-CCTTACCCAGATCCATCAAG-3'
S/L9/+15288	5'-CGCAAACATAAACAATTGCTG-3'
S/L10/-16914	5'-AGTGTTGGGTACAAGCCAGT-3'
S/L10/-17466	5'-GTTCCAAGGAACATGTCTGG-3'
S/L10/-18022	5'-AGGTGCCTGTGTAGGATGAA-3'
S/L10/+18245	5'-GGGCTGTCATGCAACTAGAG-3'
S/L10/+17663	5'-TCTTACACGCAATCCTGCTT-3'

S/L10/+17061	5'-TACCCATCTGCTCGCATAGT-3'
S/L11/-18877	5'-GCAAGCAGAATTAACCCTCA-3'
S/L11/-19396	5'-AGCACCACCTAAATTGCATC-3'
S/L11/-20002	5'-TGGTCCCTTTGAAGGTGTTA-3'
S/L11/+20245	5'-TCGAACACATCGTTTATGGA-3'
S/L11/+19611	5'-GAAGCACCTGTTTCCATCAT-3'
S/L11/+19021	5'-ACGATGCTCAGCCATGTAGT-3'
SARS/L1/F3/+800	5'-GAGGTGCAGTCACTCGCTAT-3'
SARS/L1/F4/+1391	5'-CAGAGATTGGACCTGAGCAT-3'
SARS/L1/F5/+1925	5'-CAGCAAACCACTCAATTCCT-3'
SARS/L1/R3/-1674	5'-AAATGATGGCAACCTCTTCA-3'
SARS/L1/R4/-1107	5'-CACGTGGTTGAATGACTTTG-3'
SARS/L1/R5/-520	5'-ATTTCTGCAACCAGCTCAAC-3'
SARS/L2/F3/+2664	5'-CGCATTGTCTCCTGGTTTAC-3'
SARS/L2/F4/+3232	5'-GAGATTGAGCCAGAACCAGA-3'
SARS/L2/F5/+3746	5'-ATGAGCAGGTTGTCATGGAT-3'
SARS/L2/R3/-3579	5'-CTGCCTTAAGAAGCTGGATG-3'
SARS/L2/R4/-2991	5'-TTTCTTCACCAGCATCATCA-3'
SARS/L2/R5/-2529	5'-CACCGTTCTTGAGAACAACC-3'
SARS/L3/F3/+4708	5'-TCTTTGGCTGGCTCTTACAG-3'
SARS/L3/F4/+5305	5'-GCTGGTGATGCTGCTAACTT-3'
SARS/L3/F5/+5822	5'-CCATCAAGCCTGTGTCGTAT-3'
SARS/L3/R3/-5610	5'-CAGGTGGTGCAGACATCATA-3'
SARS/L3/R4/-4988	5'-AACATCAGCACCATCCAAGT-3'
SARS/L3/R5/-4437	5'-ATCGGACACCATAGTCAACG-3'

Les séquences des fragments L0 à L12 de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, correspondent respectivement aux séquences SEQ ID NO :41 à SEQ ID NO :54 dans la liste de séquences jointe en annexe. Parmi ces séquences, seule celle correspondant aux fragments L5 comporte une différence nucléotidique par rapport à la séquence correspondante de l'isolat AY278741-Urbani. Cette mutation t/c en position 7919 aboutit à une modification de la séquence en acides aminés de la protéine correspondante, codée par l'ORF 1a: en position 2552, une valine (codon gtt ; AY278741) est changée en alanine (codon gct) dans la souche de SARS-CoV 031589. En revanche, aucune mutation n'a été identifiée par rapport à la séquence correspondante de l'isolat AY274119.3-Urbani. Les autres fragments ne présentent pas de différences par rapport aux séquences correspondantes des isolats Tor2 et Urbani.

**Exemple 2 : Production et purification de protéines N et S recombinantes de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589**

La protéine entière et deux fragments polypeptidiques de la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589 ont été produites chez *E. coli*, sous forme de protéines de fusion comprenant une étiquette polyhistidine N-ou C-terminale. Dans les deux polypeptides S, les séquences hydrophobes N et C-terminales de la protéine S (peptide signal : positions 1 à 13 et hélice transmembranaire : positions 1196 à 1218) ont été déléetées alors que l'hélice  $\beta$  (positions 565 à 687) et les deux motifs de type coiled-coils (positions 895 à 980 et 1155 à 1186) de la protéine S ont été préservés. Ces deux polypeptides sont constitués par : un fragment long ( $S_L$ ) correspondant aux positions 14 à 1193 de la séquence en acides aminés de la protéine S et un fragment court ( $S_C$ ) correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S.

**1) Clonage des ADNc N,  $S_L$  et  $S_C$  dans les vecteurs d'expression pIVEX2.3 et pIVEX2.4**

Les ADNc correspondant à la protéine N et aux fragments  $S_L$  et  $S_C$  ont été amplifiés par PCR dans des conditions standard, à l'aide de l'ADN polymérase Platinum Pfx® (INVITROGEN). Les plasmides SRAS-N et SRAS-S ont été utilisés comme matrice et les oligonucléotides suivants comme amorces :

5'-CCCATATGTCTGATAATGGACCCCAATCAAAC-3' (N sens, SEQ ID NO :55)  
 5'-CCCCCGGGTGCCTGAGTTGAATCAGCAGAAGC-3' (N antisens, SEQ ID NO :56)  
 5'-CCCATATGAGTGACCTTGACCGGTGCACCAC-3' ( $S_C$  sens, SEQ ID NO :57)  
 5'-CCCATATGAAACCTTGCACCCACCTGCTC-3' ( $S_L$  sens, SEQ ID NO :58)  
 5'-CCCCCGGGTTTAATATATTGCTCATATTTTCCC-3' ( $S_C$  et  $S_L$  antisens, SEQ ID NO :59).

Les amorces sens introduisent un site *NdeI* (souligné) alors que les amorces antisens introduisent un site *XmaI* ou *SmaI* (souligné). Les 3 produits d'amplification ont été purifiés sur colonne (kit *QIAquick PCR Purification*, QIAGEN) et clonés dans un vecteur approprié. L'ADN plasmidique purifié des 3 constructions (kit *QIAfilter Midi Plasmid*, QIAGEN) a été vérifié par séquençage et digéré par les enzymes *NdeI* et *XmaI*. Les 3 fragments correspondants aux ADNc N,  $S_L$  et  $S_C$  ont été purifiés sur gel d'agarose puis insérés dans les plasmides pIVEX2.3MCS (étiquette

polyhistidine C-terminale) et pIVEX2.4d (étiquette polyhistidine N-terminale) préalablement digérés par les mêmes enzymes. Après vérification des constructions, les 6 vecteurs d'expressions ainsi obtenus (pIV2.3N, pIV2.3S<sub>C</sub>, pIV2.3S<sub>L</sub>, pIV2.4N, pIV2.4S<sub>C</sub> également dénommé pIV2.4S<sub>I</sub>, pIV2.4S<sub>L</sub>) ont été ensuite utilisés, d'une part pour tester l'expression des protéines *in-vitro*, et d'autre part pour transformer la souche bactérienne BL21(DE3)pDIA17 (NOVAGEN). Ces constructions codent pour des protéines dont la masse moléculaire attendue est la suivante : pIV2.3N (47174 Da), pIV2.3S<sub>C</sub> (82897 Da), pIV2.3S<sub>L</sub> (132056 Da), pIV2.4N (48996 Da), pIV2.4S<sub>I</sub> (81076 Da) et pIV2.4S<sub>L</sub> (133877 Da).

## 10 2) Analyse de l'expression des protéines recombinantes *in-vitro* et *in vivo*

L'expression de protéines recombinantes à partir des 6 vecteurs recombinants a été testée, dans un premier temps, dans un système *in-vitro* (RTS100, Roche). Les protéines produites *in vitro*, après une incubation des vecteurs recombinants pIVEX, 4h à 30°C, dans le système RTS100, ont été analysées par western-blot à l'aide d'un anticorps anti-(his)<sub>6</sub> couplé à la peroxydase. Le résultat d'expression *in-vitro* (Figure 1) montre que seule la protéine N est exprimée en quantités importantes, cela quelle que soit la position, N- ou C-terminale, de l'étiquette polyhistidine. Dans une seconde étape, l'expression des protéines N et S a été testée *in-vivo* à 30°C dans du milieu LB, en présence ou en l'absence d'inducteur (IPTG 1mM). La protéine N est très bien produite dans ce système bactérien (Figure 2) et se retrouve principalement dans une fraction soluble après lyse des bactéries. En revanche, la version longue de S (S<sub>L</sub>) est très peu produite et complètement insoluble (Figure 3). La version courte (S<sub>C</sub>) présente également une très faible solubilité, mais un taux d'expression beaucoup plus élevé que celui de la version longue. Par ailleurs, la construction S<sub>C</sub> fusionnée à une étiquette polyhistidine en position C-terminale présente une taille plus faible que celle attendue. Une expérience d'immunodétection avec un anticorps anti-polyhistidine a montré que cette construction était incomplète. En conclusion, les deux constructions, pIV2.3N et pIV2.4S<sub>I</sub>, exprimant respectivement la protéine N entière fusionnée à l'étiquette polyhistidine en C-terminal et la protéine S courte fusionnée à l'étiquette polyhistidine en N-terminal, ont été retenues pour produire les deux protéines en grande quantité afin de les purifier.

### 3) Analyse de l'activité antigénique des protéines recombinantes

L'activité antigénique des protéines N, S<sub>L</sub> et S<sub>C</sub> a été testée par western-blot, à l'aide de deux échantillons de sérum, provenant d'un même patient infecté par le SARS-CoV, prélevés 8 jours (M12) et 29 jours-(M13) après le début des symptômes du SRAS. Le protocole expérimental est comme décrit à l'exemple 3. Les résultats illustrés par la figure 4 montrent (i) la séroconversion du patient, et (ii) que la protéine N possède une plus forte réactivité antigénique que la protéine S courte.

### 4) Purification de la protéine N à partir de pIV2.3N

Plusieurs expériences de purification de la protéine N, produite à partir du vecteur pIV2.3N, ont été réalisées selon le protocole suivant. Les bactéries BL21(DE3)pDIA17, transformées par le vecteur d'expression pIV2.3N, ont été cultivées à 30°C dans 1 litre de milieu de culture contenant 0,1 mg/ml d'ampicilline, et induites par 1 mM IPTG quand la densité cellulaire, équivalente à A<sub>600</sub> = 0,8, est atteinte (environ 3 heures). Après 2 heures de culture en présence d'inducteur, les cellules ont été récupérées par centrifugation (10 min à 5000 rpm), remises en suspension dans le tampon de lyse (50 mM NaH<sub>2</sub>PO<sub>4</sub>, NaCl 0,3 M, 20 mM imidazole, pH 8 contenant le mélange d'inhibiteurs de protéases *Complete*®, Roche), et lysées par la presse de French (12000 psi). Après centrifugation du lysat bactérien (15 min à 12000 rpm), le surnageant (50 ml) a été déposé à un débit de 1ml/min sur une colonne (15 ml) de chélation métallique (Ni-NTA superflow, Qiagen), équilibrée par le tampon de lyse. Après lavage de la colonne par 200 ml de tampon de lyse, la protéine N a été éluée par un gradient d'imidazole (20 → 250 mM) en 10 volumes de colonne. Les fractions contenant la protéine N ont été rassemblées et analysées par électrophorèse en gel de polyacrylamide en conditions dénaturantes puis coloration au bleu de Coomassie. Les résultats illustrés par la figure 5 montrent que le protocole employé permet de purifier la protéine N avec une homogénéité très satisfaisante (95%) et un rendement moyen de 15 mg de protéine par litre de culture.

### 5) Purification de la protéine S<sub>C</sub> à partir de pIV2.4S<sub>C</sub> (pIV2.4S<sub>1</sub>)

Le protocole suivi pour purifier la protéine S courte est très différent de celui décrit ci-dessus car la protéine est fortement aggrégée dans le système bactérien (corps d'inclusion). Les bactéries BL21(DE3)pDIA17, transformées par le vecteur d'expression pIV2.4S<sub>1</sub>, ont été cultivées à 30°C dans 1 litre de milieu de

culture contenant 0,1 mg/ml d'ampicilline, et induites par 1 mM IPTG quand la densité cellulaire, équivalente à  $A_{600} = 0,8$ , est atteinte (environ 3 heures). Après 2 heures de culture en présence d'inducteur, les cellules ont été récupérées par centrifugation (10 min à 5000 rpm), remises en suspension dans le tampon de lyse (0,1 M Tris-HCl, EDTA 1 mM, pH 7,5), et lysées par la presse de French (1200 psi). Après 5 centrifugation du lysat bactérien (15 min à 12000 rpm), le culot a été remis en suspension dans 25 ml de tampon de lyse contenant 2% Triton X100 et 10 mM  $\beta$ -mercaptoéthanol, puis centrifugé pendant 20 min à 12000 rpm. Le culot a été remis en suspension dans un tampon Tris-HCl 10 mM contenant 7 M urée, et mis en agitation 10 douce pendant 30 min à température ambiante. Ce dernier lavage des corps d'inclusion avec 7 M urée est nécessaire pour éliminer la plupart des protéines membranaires d'*E. coli* qui co-sédimentent avec la protéine  $S_C$  agrégée. Après une dernière centrifugation pendant 20 min à 12000 rpm, le culot final est remis en suspension dans le tampon Tris-HCl 10 mM. L'analyse électrophorétique de cette 15 préparation (Figure 6) montre que la protéine S courte peut être purifiée avec une homogénéité satisfaisante (environ 90%) à partir des corps d'inclusion (extrait insoluble).

### **Exemple 3 : Immunodominance de la protéine N**

La réactivité des anticorps présents dans le sérum des patients 20 atteints de pneumopathie atypique causée par le coronavirus associé au SRAS (SARS-CoV), vis-à-vis des différentes protéines de ce virus, a été analysée par western-blot dans les conditions décrites ci-après.

#### **1) Matériel**

##### **a) lysat de cellules infectées par le SARS-CoV**

25 Des cellules Vero E6 ( $2 \times 10^6$ ) ont été infectées par le SARS-CoV (isolat répertorié sous le numéro FFM/MA104) à une multiplicité d'infection (M.O.I.) de  $10^{-1}$  ou  $10^{-2}$  puis incubées dans du milieu DMEM contenant 2% de SVF, à 35°C dans une atmosphère contenant 5% de  $CO_2$ . 48 heures plus tard, le tapis cellulaire a été lavé avec du PBS puis lysé avec 500  $\mu$ l de tampon de dépôt préparé selon Laemmli 30 et contenant du  $\beta$ -mercaptoéthanol. Les échantillons ont ensuite été bouillis 10 minutes puis soniqués 3 fois 20 secondes.



b) anticorpsb<sub>1</sub>) sérum de patient atteint de pneumopathie atypique

Le sérum référencé au Centre National de Référence des virus  
influenzae (Région-Nord) sous le N° 20033168 est celui d'un patient français atteint  
5 d'une pneumopathie atypique causée par le SARS-CoV prélevé au jour 38 après le  
début des symptômes ; le diagnostic d'infection par le SARS-CoV a été réalisé par  
RT-PCR nichée et PCR quantitative.

b<sub>2</sub>) sérums polyclonaux de lapin monospécifiques dirigés contre la protéine N ou la protéine S

10 Les sérums sont ceux produits à partir des protéines recombinantes N  
et S<sub>C</sub> (exemple 2), selon le protocole d'immunisation décrit à l'exemple 4 ; il s'agit du  
sérum du lapin P13097 (sérum anti-N) et du sérum du lapin P11135 (sérum anti-S).

2) Méthode

20 µl de lysat de cellules infectées par le SARS-CoV à des M.O.I. de  
15  $10^{-1}$  et  $10^{-2}$  et, à titre de contrôle, 20 µl d'un lysat de cellules non infectées (mock) ont  
été séparés sur un gel SDS à 10% de polyacrylamide puis transférés sur une  
membrane de nitrocellulose. Après blocage dans une solution de PBS/lait 5%/Tween  
0,1% et lavage en PBS/Tween 0,1%, cette membrane a été hybridée pendant une nuit  
à 4°C avec : (i) l'immun-sérum N° 20033168 dilué au 1/300, 1/1000 et 1/3000 dans le  
20 tampon PBS/BSA 1%/Tween 0,1%, (ii) le sérum du lapin P13097 (sérum anti-N)  
dilué au 1/50000 dans le même tampon et (iii) le sérum du lapin P11135 (sérum anti-  
S) dilué au 1/10000 dans le même tampon. Après lavage en PBS/Tween, une hybrida-  
tion secondaire a été réalisée à l'aide, soit d'anticorps polyclonaux de mouton dirigés  
contre les chaînes lourdes et légères des immunoglobulines G humaines et couplés à la  
25 peroxidase (NA933V, Amersham), soit d'anticorps polyclonaux d'âne dirigés contre  
les chaînes lourdes et légères des immunoglobulines G de lapin et couplés à la  
peroxidase (NA934V, Amersham). Les anticorps fixés ont été révélés à l'aide du kit  
ECL+ (Amersham) et de films d'autoradiographie Hyperfilm MP (Amersham). Une  
échelle de masse moléculaire (kDa) est portée sur la figure.

### 3) Résultats

La figure 7 montre que trois polypeptides de masse moléculaire apparente 35, 55 et 200 kDa sont détectés spécifiquement dans les extraits de cellules infectées par le SARS-CoV.

5 Afin d'identifier ces polypeptides, deux autres immunoempreintes (figure 8) ont été réalisées sur les mêmes échantillons et dans les mêmes conditions avec des anticorps polyclonaux de lapins spécifique de la nucléoprotéine N (lapin P13097, figure 8A) et de la protéine de spicule S (lapin P11135, figure 8B) Cette expérience montre que le polypeptide de 200 kDa correspond à la glycoprotéine de  
 10 spicule S du SARS-CoV, que le polypeptide de 55 kDa correspond à la nucléoprotéine N tandis que le polypeptide de 35 kDa représente vraisemblablement une forme tronquée ou dégradée de la N.

Les données présentées dans la figure 7 montrent donc que le sérum 20033168 réagit fortement avec la N et beaucoup plus faiblement avec la S du SARS-  
 15 CoV, puisque les polypeptides de 35 et 55 kDa sont révélés sous la forme de bandes intenses pour des dilutions de 1/300, 1/1000 et 1/3000 de l'immunsérum alors que le polypeptide de 200 kDa n'est que faiblement révélé pour une dilution de 1/300. On peut noter également qu'aucun autre polypeptide du SARS-CoV n'est détecté pour des dilutions supérieures au 1/300 du sérum 20033168.

20 Cette expérience indique que la réponse en anticorps spécifique de la N du SARS-CoV domine les réponses en anticorps spécifiques des autres polypeptides du SARS-CoV et en particulier la réponse en anticorps dirigée contre la glycoprotéine S. Elle indique une immunodominance de la nucléoprotéine N lors des infections humaines par le SARS-CoV.

#### 25 **Exemple 4 : Préparation d'anticorps polyclonaux monospécifiques dirigés contre les protéines N et S du coronavirus associé au SRAS (SARS-CoV)**

##### **1) Matériel et méthode**

Trois lapins (P13097, P13081, P13031) ont été immunisés avec le polypeptide recombinant purifié correspondant à l'intégralité de la nucléoprotéine (N),  
 30 préparé selon le protocole décrit à l'exemple 2. Après une première injection de 0,35 mg par lapin de protéine émulsionnée en adjuvant complet de Freund (voie intradermique), les animaux ont reçus 3 injections de rappel à 3 puis 4 semaines

d'intervalle, de 0,35 mg de protéine recombinante émulsionnée en adjuvant incomplet de Freund.

Trois lapins (P11135, P13042, P14001) ont été immunisés avec le polypeptide recombinant correspondant au fragment court de la protéine S (S<sub>C</sub>),  
5 produit comme décrit à l'exemple 2. Comme ce polypeptide est retrouvé principalement sous la forme de corps d'inclusion dans le cytoplasme bactérien, les animaux ont reçus 4 injections intra-dermiques à 3-4 semaines d'intervalle d'une préparation de corps d'inclusion correspondant à 0,5 mg de protéine recombinante émulsionnée en adjuvant incomplet de Freund. Les 3 premières injections ont été réalisées avec une  
10 préparation de corps d'inclusion préparés selon le protocole décrit à l'exemple 2, tandis que la quatrième injection a été réalisée avec une préparation de corps d'inclusion qui ont été préparés selon le protocole décrit à l'exemple 2 puis purifiés sur gradient de saccharose et lavés en 2 % Triton X100.

Pour chaque lapin, un sérum pré-immun (p.i.) a été préparé avant la  
15 première immunisation et un immun-sérum (I.S.) 5 semaines après la quatrième immunisation.

Dans un premier temps, la réactivité des sérums a été analysée par test ELISA vis à vis de préparations de protéines recombinantes semblables à celles utilisées pour les immunisations ; les tests ELISA ont été réalisés selon le protocole et  
20 avec les réactifs tels que décrits à l'exemple 6.

Dans un deuxième temps, la réactivité des sérums a été analysée en réalisant une immunoempreinte (western blot) d'un lysat de cellules infectées par le SARS-CoV, en suivant le protocole tel que décrit à l'exemple 3.

## 2) Résultats

25 Les tests ELISA (figure 9) démontrent que les préparations de protéine N recombinante et de corps d'inclusion du fragment court de la protéine S (S<sub>C</sub>) sont immunogènes chez l'animal et que le titre des sérums immuns est élevé (plus de 1/25000).

L'immunoempreinte (figure 8) montre que le sérum immun du lapin  
30 P13097 reconnaît deux polypeptides présents dans les lysats de cellules infectées par le SARS-CoV : un polypeptide dont la masse moléculaire apparente (50-55 kDa selon les expériences) est compatible avec celle de la nucléoprotéine N (422 résidus, masse

moléculaire prédite de 46 kDa) et un polypeptide de 35 kDa, qui représente vraisemblablement une forme tronquée ou dégradée de la N.

Cette expérience montre également que le sérum du lapin P11135 reconnaît principalement un polypeptide dont la masse moléculaire apparente (180-  
 5 220 kDa selon les expériences) est compatible avec une forme glycosylée de la S (1255 résidus, chaîne polypeptidique non glycosylée de 139 kDa), ainsi que des polypeptides plus légers, qui représentent vraisemblablement des formes tronquées et/ou non glycosylées de la S.

En conclusion, l'ensemble de ces expériences démontrent que des  
 10 polypeptides recombinants exprimés chez *E. coli* et correspondant aux protéines N et S du SARS-CoV permettent d'induire chez l'animal des anticorps polyclonaux capables de reconnaître les formes natives de ces protéines.

**Exemple 5 : Préparation d'anticorps polyclonaux monospécifiques dirigés contre les protéines M et E du coronavirus associé au SRAS (SARS-CoV)**

15 **1) Analyse de la structure des protéines M et E**

a) Protéine E

La structure de la protéine E du SARS-CoV (76 acides aminés) a été analysée *in silico*, à l'aide de différents logiciels comme signalP v1.1, NetNGlyc 1.0, THMM 1.0 et 2.0 (Krogh et al., 2001, J. Mol. Biol., 305(3):567-580) ou encore  
 20 TOPPRED (von Heijne, 1992, J. Mol. Biol. 225, 487-494). L'analyse montre que ce polypeptide non glycosylé est une protéine membranaire de type 1, contenant une seule hélice transmembranaire (aa 12-34 d'après THMM), et dont la plus grande partie du domaine hydrophile (42 résidus) est localisée à l'extrémité C-terminale et vraisemblablement à l'intérieur de la particule virale (endodomaine). On peut noter une inversion  
 25 dans la topologie prédite par les versions 1.0 (N-ter est externe) et 2.0 (N-ter est interne) du logiciel THMM, mais que d'autres algorithmes, notamment TOPPRED et THUMBUP (Zhou et Zhou, 2003, Protein Science 12 :1547-1555) confirment une localisation externe de l'extrémité N-terminale de E.

b) Protéine M

30 Une analyse similaire réalisée sur la protéine M du SARS-CoV (221 acides aminés) montre que ce polypeptide ne possède pas de peptide signal (d'après le logiciel signalP v1.1) mais trois domaines transmembranaires (résidus 15-37, 50-72,

77-99 d'après THMM2.0) et un grand domaine hydrophile (aa 100-221) localisé à l'intérieur de la particule virale (endodomaine). Elle est vraisemblablement glycosylée sur l'asparagine en position 4 (d'après NetNGlyc 1.0).

Ainsi, en accord avec les données expérimentales connues pour les autres coronavirus, il est remarquable que les deux protéines M et E présentent des endodomains correspondant à la majeure partie des polypeptides et des ectodomains de très petite taille.

- l'ectodomaine de E correspond vraisemblablement aux résidus 1 à 11 ou 1 à 12 de la protéine : MYSFVSEETGT(L), SEQ ID NO : 70. En effet, la probabilité associée à la localisation transmembranaire du résidu 12 est intermédiaire (0,56 d'après THMM 2.0).

- l'ectodomaine de M correspond vraisemblablement aux résidus 2 à 14 de la protéine : ADNGTITVEELKQ, SEQ ID NO : 69. En effet, la méthionine N-terminale de M est très probablement clivée du polypeptide mature car le résidu en position 2 est une Alanine (Varshavsky, 1996, 93:12142-12149).

Par ailleurs, l'analyse de l'hydrophobicité (Kyte & Doolittle, Hopp & Woods) de la protéine E met en évidence que l'extrémité C-terminale de l'endodomaine de E est hydrophile et donc vraisemblablement exposée à la surface de ce domaine. Ainsi, un peptide synthétique correspondant à cette extrémité est un bon candidat immunogène pour induire chez l'animal des anticorps dirigés contre l'endodomaine de E. En conséquence, un peptide correspondant aux 24 résidus C-terminaux de E a été synthétisé.

## **2) Préparation d'anticorps dirigés contre l'ectodomaine des protéines M et E et l'endodomaine de la protéine E**

Les peptides M2-14 (ADNGTITVEELKQ, SEQ ID NO : 69), E1-12 (MYSFVSEETGTL, SEQ ID NO : 70) et E53-76 (KPTVYVYSRVKNLNSSEGVP DLLV, SEQ ID NO : 71) ont été synthétisés par Neosystem. Ils ont été couplés à la KLH (*Keyhole Limpet Hemocyanin*) à l'aide du MBS (m-maleimido-benzoyl-N-hydroxysuccinimide ester) via une cystéine ajoutée au cours de la synthèse soit en N-terminal du peptide (cas de E53-76) soit en C-terminal (cas de M2-14 et E1-12).

Deux lapins ont été immunisés avec chacun des conjugués, en suivant le protocole d'immunisation suivant : après une première injection de 0,5 mg de peptide couplé à la KLH et émulsionné en adjuvant complet de Freund (voie intradermique), les animaux reçoivent 2 à 4 injections de rappel à 3 ou 4 semaines d'intervalle de 0,25 mg de peptide couplé à la KLH et émulsionné en adjuvant incomplet de Freund.

Pour chaque lapin, un sérum pré-immun (p.i.) a été préparé avant la première immunisation et un immun-sérum (I.S.) est préparé 3 à 5 semaines après les injections de rappel.

La réactivité des sérums est analysée dans un premier temps par test ELISA vis à vis du peptide utilisé pour l'immunisation, puis par immunoempreinte vis-à-vis de lysats de cellules infectées par le SARS-CoV, comme décrit pour les sérums anti-N et anti-S de l'exemple 4, selon des protocoles similaires à ceux décrits aux exemples 3 et 6, respectivement pour l'immunoempreinte et le test ELISA.

Dans un second temps, la réactivité des immunosérums dirigés contre les peptides M2-14 et E1-12 à reconnaître les ectodomains de M et de E présents à la surface de la particule virale native est analysée par des tests d'immunocapture et/ou d'immunoprécipitation de virions natifs.

**Exemple 6 : Analyse de la réactivité en ELISA de la protéine N recombinante, vis-à-vis de sérums de patients atteints de SRAS**

### 1) Matériel

L'antigène utilisé pour préparer les phases solides est la nucléoprotéine N recombinante purifiée préparée selon le protocole décrit à l'exemple 2.

Les sérums à tester (Tableau IV) ont été choisis sur la base des résultats d'analyse de leur réactivité par immunofluorescence (titre IF-SRAS), vis-à-vis de lysats de cellules infectées par le SARS-CoV.

**Tableau IV: Sérums testés en ELISA**

Référence	N° sérum	Type de sérum	Date du Sérum***	Titre IF-SRAS
3050	A	Témoin	na*	nt**
3048	B	Témoin	na	nt
033168	D	Patient 1- SRAS	27/04/03 (J38)	320
033397	E	Patient-1 SRAS	11/05/03 (J52)	320
032632	F	Patient-2 SRAS	21/03/03 (J17)	2500
032791	G	Patient-3 SRAS	04/04/03 (J3)	<40
033258	H	Patient-3 SRAS	28/04/03 (J27)	160

\*na : non-applicable. \*\* nt : non-testé. \*\*\* les dates indiquées correspondent au nombre de jours après le début des symptômes de SRAS.

5

## 2) Méthode

La protéine N (100 µl) diluée à différentes concentrations dans du tampon carbonate 0,1 M, pH 9,6 (1, 2 ou 4 µg/ml) est distribuée dans les puits de plaques ELISA, puis les plaques sont incubées une nuit à température du laboratoire.

- 10 Les plaques sont lavées avec du tampon PBS-Tween, saturées avec du tampon PBS-lait écrémé-saccharose (5 %). Les sérums à tester (100 µl) préalablement dilués (1/50, 1/100, 1/200, 1/400, 1/800, 1/1600 et 1/3200) sont ajoutés, puis les plaques sont incubées 1 h à 37° C. Après 3 lavages, le conjugué anti-IgG humaines marqué à la peroxydase (référence 209-035-098, JACKSON) dilué au 1/18000 est ajouté puis les plaques
- 15 sont incubées 1h à 37 °C. Après 4 lavages, le chromogène (TMB) et le substrat (H<sub>2</sub>O<sub>2</sub>) sont ajoutés et les plaques sont incubées 30min à température ambiante, à l'abri de la lumière. La réaction est ensuite arrêtée puis l'absorbance à 450 nm est mesurée à l'aide d'un lecteur automatique.

## 3) Résultats

- 20 Les tests ELISA (figure 10) démontrent que la préparation de protéine N recombinante est reconnue spécifiquement par les anticorps de sérums de patients atteints de SRAS prélevés en phase tardive de l'infection (≥ 17 jours après le début des symptômes) alors qu'elle n'est pas reconnue de façon significative par les anticorps d'un sérum de patient prélevé en phase précoce de l'infection (3 jours après
- 25 le début des symptômes) ni par des sérums témoins de sujets non atteints de SRAS.

**Exemple 7 : Détection du coronavirus associé au SRAS (SARS-CoV) par RT-PCR en temps réel à l'aide d'amorces spécifiques du gène de la nucléoprotéine**

**1) Mise au point des conditions de la RT-PCR**

**a) conception des amorces et des sondes**

- 5                    La conception des amorces et sondes a été réalisée à partir de la séquence du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589, à l'aide du programme "Light Cycler Probe Design (Roche)". Ainsi les deux séries d'amorces et de sondes suivantes ont été sélectionnées :

**- série 1 (SEQ ID NO : 60, 61, 64, 65):**

- 10 - amorce sens : N/+28507 : 5'-GGC ATC GTA TGG GTT G-3' [28507-28522]  
 - amorce antisens : N/-28774 : 5'-CAG TTT CAC CAC CTC C-3' [28774-28759]  
 - sonde 1 : 5'-GGC ACC CGC AAT CCT AAT AAC AAT GC-fluorescéine 3' [28561-28586]  
 - sonde 2 : 5' Red705 -GCC ACC GTG CTA CAA CTT CCT-phosphate [28588-28608]

15                    **- série 2 (SEQ ID NO : 62, 63, 66, 67)**

- amorce sens : N/+28375 : 5'-GGC TAC TAC CGA AGA G-3' [28375-28390]  
 - amorce antisens : N/-28702 : 5'-AAT TAC CGC GAC TAC G-3' [28702-28687]  
 - sonde 1 : SRAS/N/FL : 5'-ATA CAC CCA AAG ACC ACA TTG GC - fluorescéine 3' [28541-28563]  
 20 - sonde 2 : SRAS/N/LC705 : 5' Red705 -CCC GCA ATC CTA ATA ACA ATG CTG C-phosphate 3' [28565-28589]

**b) analyse de l'efficacité des deux couples amorces**

- 25                    Afin de tester l'efficacité respective des deux couples d'amorces, une amplification par RT-PCR a été réalisée sur un ARN synthétique correspondant aux nucléotides 28054-29430 du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589 et contenant la séquence du gène N.

De manière plus précise :

- 30                    Cet ARN synthétique a été préparé par transcription *in vitro* à l'aide de l'ARN polymérase du phage T7, d'une matrice d'ADN obtenu par linéarisation du plasmide SRAS-N avec l'enzyme *Bam HI*. Après élimination de la matrice d'ADN par digestion à l'aide de DNase 1, les ARN synthétiques sont purifiés par une extraction au phénol-chloroforme suivie de deux précipitations successives en acétate d'ammonium et isopropanol. Ils sont alors quantifiés par mesure de l'absorbance à 260



nm et leur qualité est contrôlée par le rapport des absorbances à 260 et 280 nm ainsi que par une électrophorèse en gel d'agarose. Ainsi, la concentration de la préparation d'ARN synthétique utilisée pour ces études est de 1,6 mg/ml, ce qui correspond à  $2,1.10^{15}$  copies/ml d'ARN.

5 Des quantités décroissantes d'ARN synthétique ont été amplifiées par RT-PCR à l'aide du kit "Superscript™ One-Step RT-PCR with Platinum® Taq" et les couples d'amorces n° 1 (N+/28507, N-/28774) (figure 1A) et n° 2 (N+/28375, N-/28702) (figure 1B), en suivant les indications du fournisseur. Les conditions d'amplification utilisées sont les suivantes : l'ADNc a été synthétisé par incubation 30 min à 10 45 °C, 15 min à 55°C puis 2 min à 94 °C puis il a été amplifié par 5 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec, suivis de 35 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 15 sec, avec 2 sec d'élongation supplémentaire à chaque cycle, et d'une étape finale d'élongation à 72°C pendant 5 min. Les produits d'amplification obtenus ont ensuite été maintenus à 10°C.

Les résultats présentés à la figure 11 montrent que le couple d'amorces n° 2 (N+/28375, N-/28702) permet de détecter jusqu'à 10 copies d'ARN 20 (bande de faible intensité) ou  $10^2$  copies (bande de bonne intensité) contre  $10^4$  copies pour le couple d'amorces n° 1 (N+/28507, N-/28774). Les amplicons sont respectivement de 268 pb (couple 1) et de 328 pb (couple 2).

#### c) mise au point de la RT-PCR en temps réel

25 Une RT-PCR en temps réel a été mise au point à l'aide du couple d'amorces n°2 et du couple de sonde constitué par SRAS/N/FL et SRAS/N/LC705 (figure 2).

L'amplification a été réalisée sur un LightCycler™ (Roche) à l'aide du kit "Light Cycler RNA Amplification Kit Hybridization Probes " (référence 2 015 145, Roche) dans les conditions optimisées suivantes. Un Mélange réactionnel contenant : H<sub>2</sub>O (6,8 µl), MgCl<sub>2</sub> 25 mM (0,8 µl, 4 µM final de Mg<sup>2+</sup>), mélange réactionnel 30 5X (4 µl), sonde SRAS/N/FL 3µM (0,5 µl, 0,075 µM final), sonde SRAS/N/LC705 3

$\mu\text{M}$  (0,5  $\mu\text{l}$ , 0,075  $\mu\text{M}$  final), amorce N/+28375 10  $\mu\text{M}$  (1  $\mu\text{l}$ , 0,5  $\mu\text{M}$  final), amorce N/-28702 10  $\mu\text{M}$  (1  $\mu\text{l}$ , 0,5  $\mu\text{M}$  final), mélange d'enzyme (0,4  $\mu\text{l}$ ) et échantillon (ARN viral, 5  $\mu\text{l}$ ) a été amplifié en suivant le programme suivant :

- Transcription inverse : 50°C      10:00min      analysis mode: none
- 5 - Dénaturation :      95°C      30sec x1      analysis mode: none
- Amplification : 95°C      2sec      ►
- 50°C      15sec      analysis mode: quantification\* ► x45
- 72°C      13sec      rampe thermique 2,0°C/sec      ►
- refroidissement : 40°C      30sec x1      analysis mode: none
- 10 \*La mesure de fluorescence se fait à la fin de l'hybridation et à chaque cycle (en mode SINGLE).

Les résultats présentés à la figure 12 montrent que cette RT-PCR en temps réel est très sensible puisqu'elle permet de détecter  $10^2$  copies d'ARN synthétique dans 100% des 5 échantillons analysés (29/29 échantillons dans 8 expériences) et jusqu'à 10 copies d'ARN dans 100% des 5 échantillons analysés (40/45 échantillons dans 8 expériences). Elle montre également que cette RT-PCR permet de détecter la présence du génome du SARS-CoV dans un échantillon et de quantifier le nombre de génomes présents. A titre d'exemple, l'ARN viral d'un stock de SARS-CoV cultivé sur cellules Vero E6 a été extrait à l'aide du kit "Qiam viral RNA extraction" (Qiagen), dilué à  $0,05 \cdot 10^{-4}$  et analysé par RT-PCR en temps réel selon le protocole décrit ci-dessus; l'analyse présentée à la figure 12 montre que ce stock de virus contient  $6,5 \cdot 10^9$  génomes –équivalents/ml (geq/ml), ce qui est tout à fait similaire à la valeur de  $1,0 \cdot 10^{10}$  geq/ml mesurée à l'aide du kit "RealArt™ HPA-Coronavirus LC RT PCR Reagents" commercialisé par Artus.

#### 25 d) détection de l'ARN du SARS-CoV par PCR en temps réel à partir de prélèvements respiratoires

Une étude comparative a été réalisée sur une série de prélèvements respiratoires reçus par le Centre National de Référence du Virus Influenzae (région nord) et susceptibles de contenir du SARS-CoV. Pour ce faire, l'ARN a été extrait des  
30 prélèvements à l'aide du kit "Qiam viral RNA extraction" (Qiagen) et analysé par RT-PCR en temps réel, d'une part à l'aide des couples d'amorces et de sondes de la

série n° 2 dans les conditions décrites ci-dessus d'une part, et d'autre part à l'aide du kit "LightCycler SARS-CoV quantification kit" commercialisé par Roche (référence 03 604 438). Les résultats sont résumés dans le Tableau ci-dessous. Ils montrent que 18 des 26 prélèvements sont négatifs et 5 des 26 prélèvements sont positifs pour les deux kits, tandis qu'un prélèvement est positif pour le seul kit Roche et deux pour les seuls réactifs N"série2". En outre, pour 3 prélèvements (20032701, 20032712, 20032714) les quantités d'ARN détectés sont nettement supérieures avec les réactifs (sondes et amorces) de la série n°2. Ces résultats indiquent que les amorces et sondes N"série2" sont plus sensibles pour la détection du génome du SARS-CoV dans des prélèvements biologiques que celles du kit actuellement disponible.

**Tableau V: Analyse par RT-PCR en temps réel des ARN extraits d'une série de prélèvements de 5 patients à l'aide des couples d'amorces et de sondes de la série n° 2 (N "série 2") ou du kit "LightCycler SARS-CoV quantification kit" (Roche). Le type de prélèvement est indiqué ainsi que le nombre de copies de génome viral mesurées dans chacun des deux tests. NEG : RT-PCR négative.**

Prélèvements n°	Patient	Type de prélèvement	KIT ROCHE	N "série2"
20033082	K	nasal	NEG	NEG
20033083	K	pharyngé	NEG	NEG
20033086	K	nasal	NEG	NEG
20033087	K	pharyngé	NEG	NEG
20032802	M	nasal	NEG	NEG
20032803	M	expectoration	NEG	NEG
20032806	M	nasal ou pharyngé	NEG	NEG
20031746ARN2	C	pharyngé	NEG	NEG
20032711	C	nasal ou pharyngé	<b>39</b>	NEG
20032910	B	nasal	NEG	NEG
20032911	B	pharyngé	NEG	NEG
20033356	V	expectoration	NEG	NEG
20033357	V	expectoration	NEG	NEG
20031725	K	asp. endotrachéale	NEG	<b>150</b>
20032657	K	asp. endotrachéale	NEG	NEG
20032698	K	asp. endotrachéale	NEG	NEG
20032720	K	asp. endotrachéale	<b>3</b>	<b>5</b>
20033074	K	selles	<b>115</b>	<b>257</b>
20032701	M	pharyngé	<b>443</b>	<b>1676</b>
20032702	M	expectoration	NEG	<b>249</b>
20031747ARN2	C	pharyngé	NEG	NEG
20032712	C	inconnu	<b>634</b>	<b>6914</b>
20032714	C	pharyngé	<b>17</b>	<b>223</b>
20032800	B	nasal	NEG	NEG
20033353	V	nasal	NEG	NEG
20033384	V	nasal	NEG	NEG

### REVENDICATIONS

1°) Utilisation d'un produit sélectionné dans le groupe constitué par :

a) une protéine ou un peptide codé par le polynucléotide de séquence

SEQ ID NO : 1,

5

b) un anticorps ou un fragment d'anticorps monoclonal ou polyclonal dirigé contre ladite protéine ou ledit peptide en a), et

c) une puce ou un filtre à protéine ou à peptide comprenant la protéine ou le peptide en a) ou bien l'anticorps ou le fragment d'anticorps en b), pour la préparation d'un réactif de détection et éventuellement de sérotypage, d'un

10 coronavirus associé au SRAS.

2°) Utilisation selon la revendication 1, caractérisée en ce que ladite protéine est sélectionnée dans le groupe constitué par :

- la protéine S de séquence SEQ ID NO : 3 ,

- la protéine E de séquence SEQ ID NO : 14,

15

- la protéine M de séquence SEQ ID NO : 17,

- la protéine N de séquence SEQ ID NO : 37, et

- les protéines codées par les ORF : ORF1a, ORF1b, ORF3, ORF4 et ORF7 à ORF11, ORF13 et ORF14 de séquence respectivement, SEQ ID NO : 74, 75, 10, 12, 22, 24, 26, 28, 30, 33 et 35 .

20

3°) Utilisation selon la revendication 1, caractérisée en ce que ledit peptide est sélectionné dans le groupe constitué par :

a) les peptides correspondant aux positions 14 à 1193 et 475 à 1193 de la séquence en acides aminés de la protéine S,

b) les peptides correspondant aux positions 2 à 14 (SEQ ID NO : 69) et 100 à 221 de la séquence en acides aminés de la protéine M ; et

c) les peptides correspondant aux positions 1 à 12 (SEQ ID NO : 70) et 53 à 76 (SEQ ID NO : 71) de la séquence en acides aminés de la protéine E ; et

les peptides de 5 à 50 acides aminés consécutifs, de préférence de 10 à 30 acides aminés, inclus ou chevauchant partiellement ou totalement la séquence des peptides

30 tels que définis en a), b) ou c).

## REVENDICATIONS

1°) Utilisation d'un produit sélectionné dans le groupe constitué par :

a) une protéine ou un peptide codé par le polynucléotide de séquence  
SEQ ID NO : 1,

5                   b) un anticorps ou un fragment d'anticorps monoclonal ou polyclonal dirigé contre ladite protéine ou ledit peptide en a), et

                  c) une puce ou un filtre à protéine ou à peptide comprenant la protéine ou le peptide en a) ou bien l'anticorps ou le fragment d'anticorps en b),  
pour la préparation d'un réactif de détection et éventuellement de sérotypage, d'un  
10 coronavirus associé au SRAS.

2°) Utilisation selon la revendication 1, caractérisée en ce que ladite protéine est sélectionnée dans le groupe constitué par :

- la protéine S de séquence SEQ ID NO : 3 ,  
- la protéine E de séquence SEQ ID NO : 14,  
15           - la protéine M de séquence SEQ ID NO : 17,  
- la protéine N de séquence SEQ ID NO : 37, et  
- les protéines codées par les ORF : ORF1a, ORF1b, ORF3, ORF4  
et ORF7 à ORF11, ORF13 et ORF14 de séquence respectivement, SEQ ID NO : 74,  
75, 10, 12, 22, 24, 26, 28, 30, 33 et 35 .

20           3°) Utilisation selon la revendication 1, caractérisée en ce que ledit peptide est sélectionné dans le groupe constitué par :

a) les peptides correspondant aux positions 14 à 1193 et 475 à 1193 de la séquence en acides aminés de la protéine S,

b) les peptides correspondant aux positions 2 à 14 (SEQ ID NO : 69)  
25 et 100 à 221 de la séquence en acides aminés de la protéine M ; et

c) les peptides correspondant aux positions 1 à 12 (SEQ ID NO : 70) et 53 à 76 (SEQ ID NO : 71) de la séquence en acides aminés de la protéine E ; et les peptides de 5 à 50 acides aminés consécutifs, de préférence de 10 à 30 acides aminés, inclus ou chevauchant partiellement ou totalement la séquence des peptides  
30 tels que définis en a), b) ou c).

4°) Utilisation selon la revendication 1, caractérisée en ce que ledit peptide est constitué de 7 à 50 acides aminés consécutifs codés par le polynucléotide

4°) Utilisation selon la revendication 1, caractérisée en ce que ledit peptide présente une séquence de 7 à 50 incluant un résidu d'acide aminé sélectionné dans le groupe constitué par :

5 - l'alanine située en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a de la souche isolée de coronavirus telle que définie à la revendication 1 ou à la revendication 2,

- la sérine située en position 577 de la séquence en acides aminés de la protéine S de la souche isolée de coronavirus telle que définie à la revendication 1 ou à la revendication 2,

10 - la glycine en position 11 de la séquence en acides aminés de l'ORF3 de la souche isolée de coronavirus telle que définie à la revendication 1 ou à la revendication 2, et

- la sérine en position 154 de la séquence en acides aminés de la protéine M de la souche isolée de coronavirus telle que définie à la revendication 1 ou à la revendication 2.

5°) Méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

(a) la mise en contact dudit échantillon biologique avec au moins un anticorps ou un fragment d'anticorps, une protéine, un peptide ou bien une puce ou un filtre à protéine ou à peptide tels que définis à l'une quelconque des revendications 1 à 4, et

(b) la révélation par tout moyen approprié des complexes antigène-anticorps formés en (a).

25 6°) Méthode selon la revendication 5, caractérisée en ce que l'étape (a) comprend :

(a<sub>1</sub>) la mise en contact dudit échantillon biologique avec au moins un premier anticorps ou fragment d'anticorps qui est fixé sur un support approprié, notamment une microplaque ,

30 (a<sub>2</sub>) le lavage de la phase solide, et

de séquence SEQ ID NO : 1, lequel peptide est sélectionné dans le groupe constitué par :

- un peptide comprenant l'alanine située en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a,
- 5        - un peptide comprenant la sérine située en position 577 de la séquence en acides aminés de la protéine S,
- un peptide comprenant la glycine en position 11 de la séquence en acides aminés de la protéine codée par l'ORF3, et
- un peptide comprenant la sérine en position 154 de la séquence en
- 10    acides aminés de la protéine M.

5°) Méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

- (a) la mise en contact dudit échantillon biologique avec au moins un
- 15 anticorps ou un fragment d'anticorps, une protéine, un peptide ou bien une puce ou un filtre à protéine ou à peptide tels que définis à l'une quelconque des revendications 1 à 4, et

- (b) la révélation par tout moyen approprié des complexes antigène-anticorps formés en (a).

- 20        6°) Méthode selon la revendication 5, caractérisée en ce que l'étape (a) comprend :

- (a<sub>1</sub>) la mise en contact dudit échantillon biologique avec au moins un premier anticorps ou fragment d'anticorps qui est fixé sur un support approprié, notamment une microplaque ,

- 25        (a<sub>2</sub>) le lavage de la phase solide, et

- (a<sub>3</sub>) l'addition d'au moins un second anticorps ou fragment d'anticorps, différent du premier, ledit anticorps ou fragment d'anticorps étant éventuellement marqué de façon appropriée.

- 30        7°) Kit ou coffret de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une protéine ou un peptide, un anticorps ou un fragment d'anticorps et

(a<sub>3</sub>) l'addition d'au moins un second anticorps ou fragment d'anticorps, différent du premier, ledit anticorps ou fragment d'anticorps étant éventuellement marqué de façon appropriée.

7°) Kit ou coffret de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une protéine ou un peptide, un anticorps ou un fragment d'anticorps et une puce ou un filtre à protéine ou à peptide tels que définis à l'une quelconque des revendications 1 à 4.

8°) Composition immunogène, caractérisée en ce qu'elle comprend au moins un produit sélectionné dans le groupe constitué par :

a) une protéine ou un peptide tels que définis à la revendication 1,  
b) un polynucléotide de type ADN ou ARN ou l'un de ses fragments représentatifs tels que définis ci-dessus, de séquence choisie parmi :

(i) la séquence SEQ ID NO : 1 ou son équivalent ARN  
(ii) la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,

(iii) la séquence complémentaire de la séquence SEQ ID NO : 1 ou de la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,

(iv) la séquence nucléotidique d'un fragment représentatif du polynucléotide tel que défini en (i), (ii) ou (iii),

(v) la séquence telle que définie en (i), (ii), (iii) ou (iv), modifiée, et  
c) un vecteur d'expression recombinant comprenant un polynucléotide tel que défini en b), et

d) une banque d'ADNc telle que définie ci-dessus.

9°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75 pour former un complexe immun avec un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

10°) Complexe immun formé d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les



une puce ou un filtre à protéine ou à peptide tels que définis à l'une quelconque des revendications 1 à 4.

8°) Composition immunogène, caractérisée en ce qu'elle comprend au moins un produit sélectionné dans le groupe constitué par :

- 5 a) une protéine ou un peptide tels que définis à la revendication 1,
- b) un polynucléotide de type ADN ou ARN ou l'un de ses fragments représentatifs, de séquence choisie parmi :
  - (i) la séquence SEQ ID NO : 1 ou son équivalent ARN
  - (ii) la séquence hybridant dans des conditions de forte stringence
  - 10 avec la séquence SEQ ID NO : 1,
  - (iii) la séquence complémentaire de la séquence SEQ ID NO : 1 ou de la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,
  - (iv) la séquence nucléotidique d'un fragment représentatif du poly-
  - 15 nucléotide tel que défini en (i), (ii) ou (iii),
  - (v) la séquence telle que définie en (i), (ii), (iii) ou (iv), modifiée, et
  - c) un vecteur d'expression recombinant comprenant un polynucléotide tel que défini en b), et
  - d) une banque d'ADNc telle que définie ci-dessus.

20 9°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, *in vitro*, pour former un complexe immunitaire avec un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

25 10°) Complexe immunitaire formé d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

30 11°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, pour la

une puce ou un filtre à protéine ou à peptide tels que définis à l'une quelconque des revendications 1 à 4.

8°) Composition immunogène, caractérisée en ce qu'elle comprend au moins un produit sélectionné dans le groupe constitué par :

- 5 a) une protéine ou un peptide tels que définis à la revendication 1,
- b) un polynucléotide de type ADN ou ARN ou l'un de ses fragments représentatifs, de séquence choisie parmi :
  - (i) la séquence SEQ ID NO : 1 ou son équivalent ARN
  - (ii) la séquence hybridant dans des conditions de forte stringence avec la
   
10 séquence SEQ ID NO : 1,
  - (iii) la séquence complémentaire de la séquence SEQ ID NO : 1 ou de la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,
  - (iv) la séquence nucléotidique d'un fragment représentatif du poly-
   
15 nucléotide tel que défini en (i), (ii) ou (iii),
  - (v) la séquence telle que définie en (i), (ii), (iii) ou (iv), modifiée, et
  - c) un vecteur d'expression recombinant comprenant un polynucléotide tel que défini en b), et
  - d) une banque d'ADNc comprenant un polynucléotide tel que défini en b).
- 20 9°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, *in vitro*, pour former un complexe immun avec un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.
- 25 10°) Complexe immun formé d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.
- 30 11°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, pour la

séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

- 5 11°) Utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75 pour induire la production d'un anticorps capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS.

- 10 12°) Utilisation d'un polynucléotide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 1, 2, 4, 7, 8, 13, 15, 16, 18, 19, 20, 31, 36 et 38 pour induire la production d'un anticorps dirigé contre la protéine codée par ledit polynucléotide et capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS

préparation d'une composition immunogène apte à induire la production d'un anticorps capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS.

- 12°) Utilisation d'un polynucléotide isolé ou purifié présentant une
- 5 séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 1, 2, 4, 7, 8, 13, 15, 16, 18, 19, 20, 31, 36 et 38, pour la préparation d'une composition immunogène apte à induire la production d'un anticorps dirigé contre la protéine codée par ledit polynucléotide et capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS.

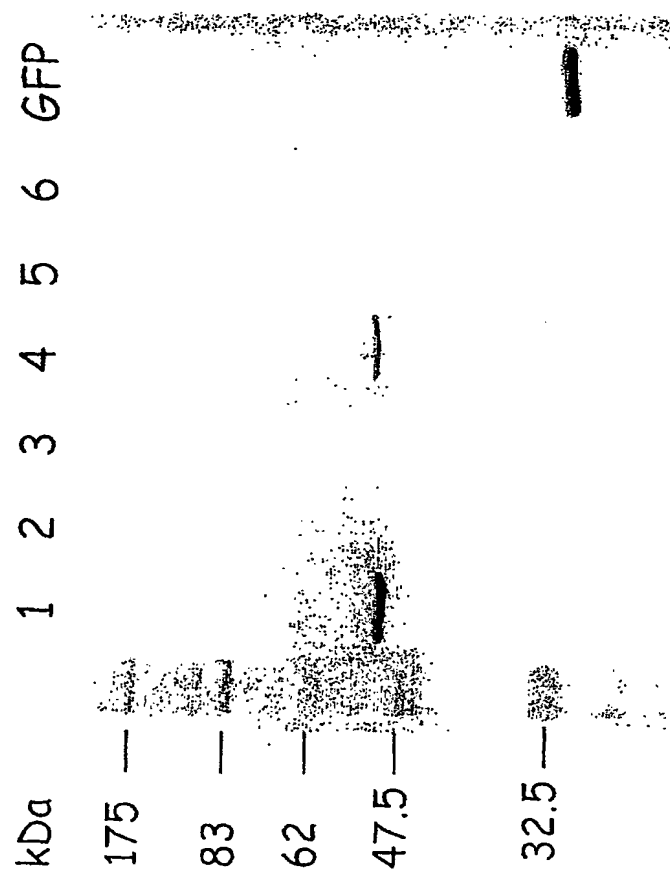


Figure 1

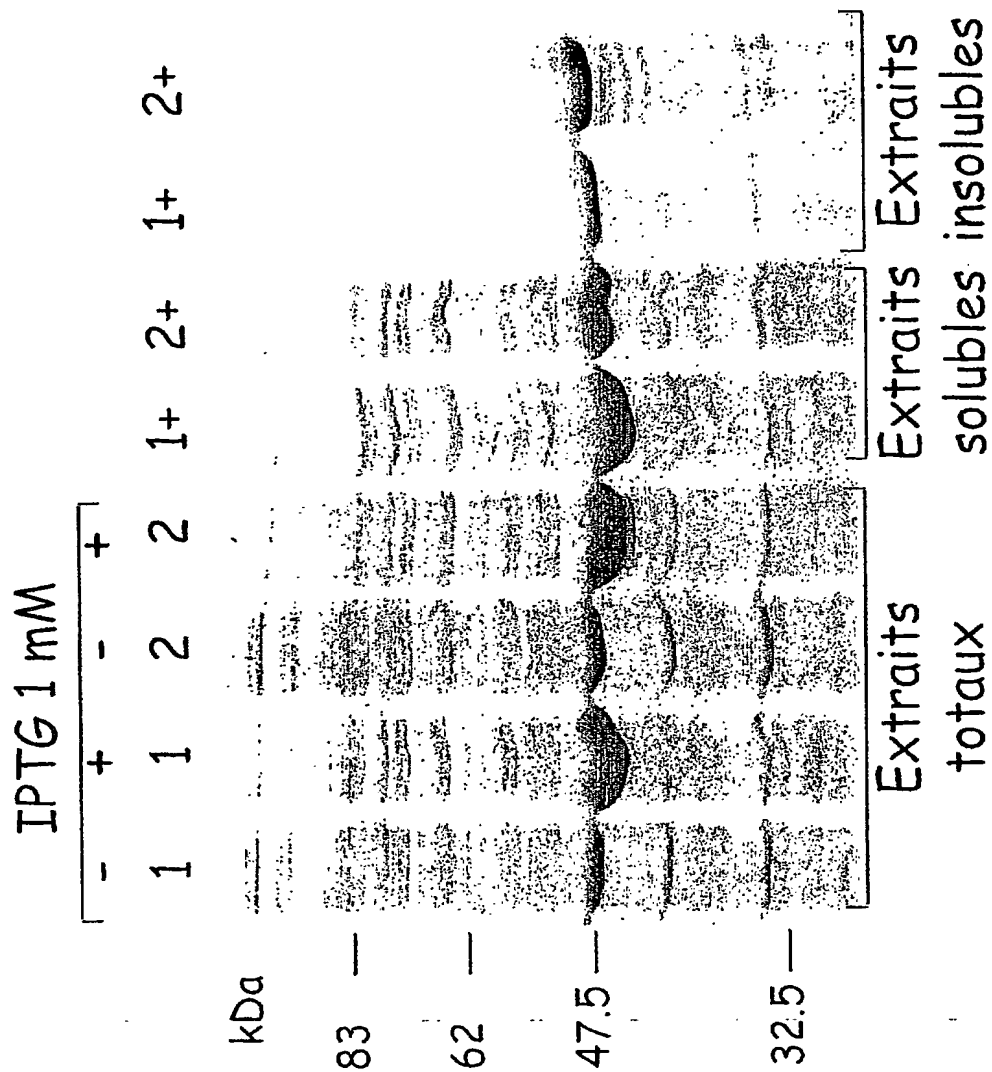


Figure 2

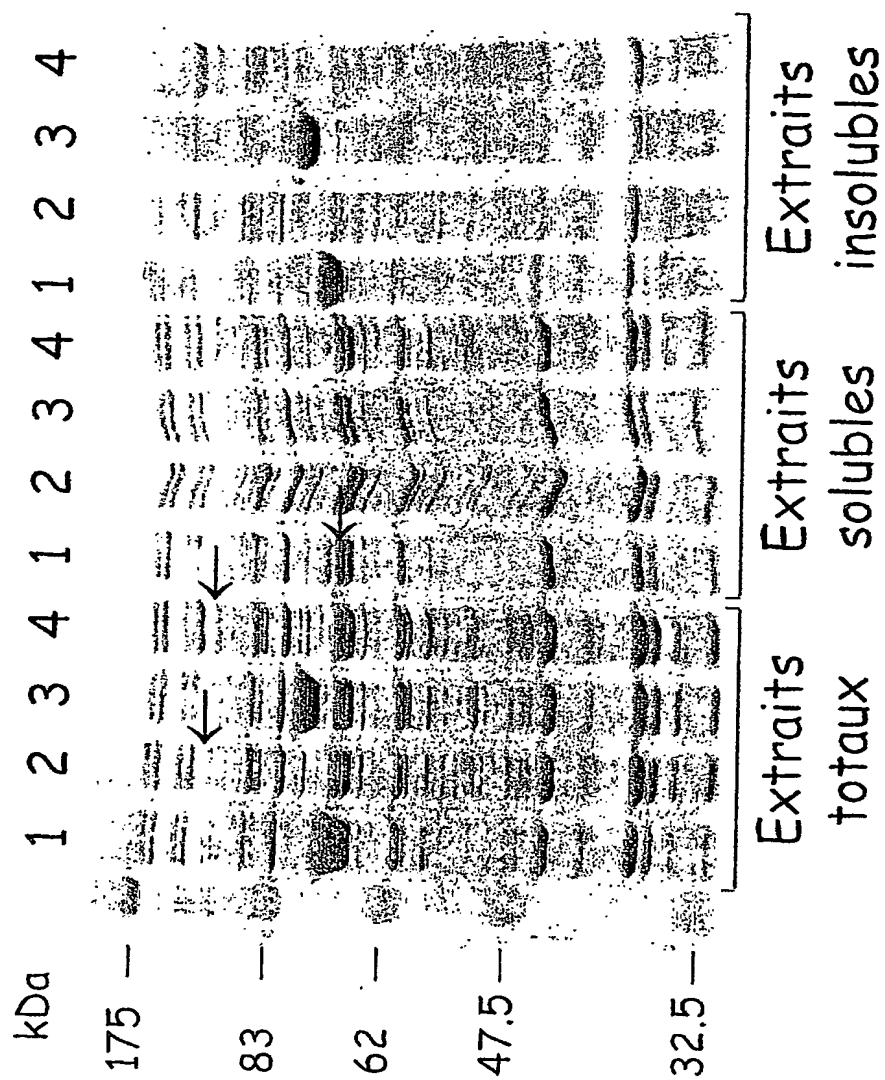


Figure 3

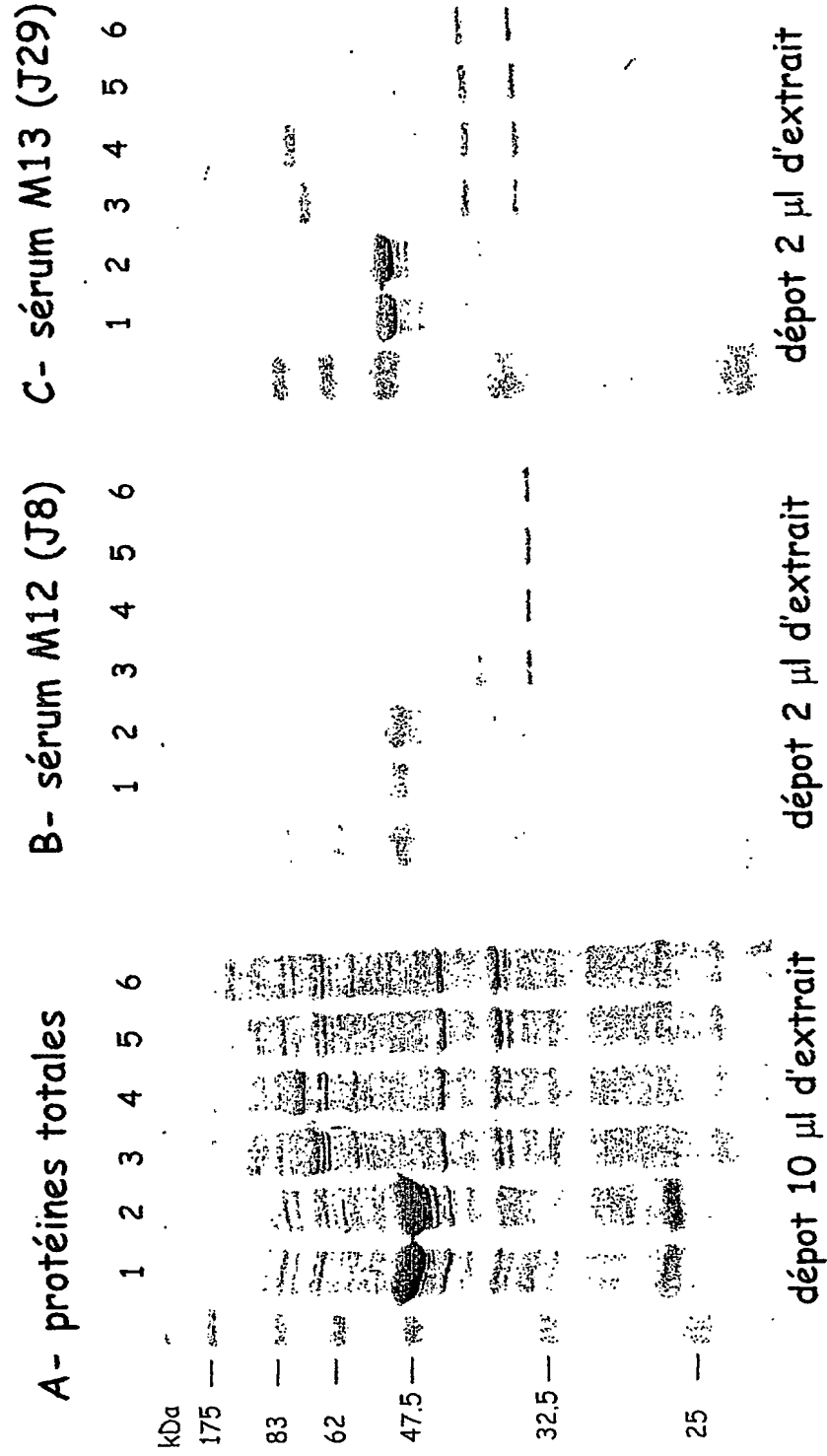


Figure 4



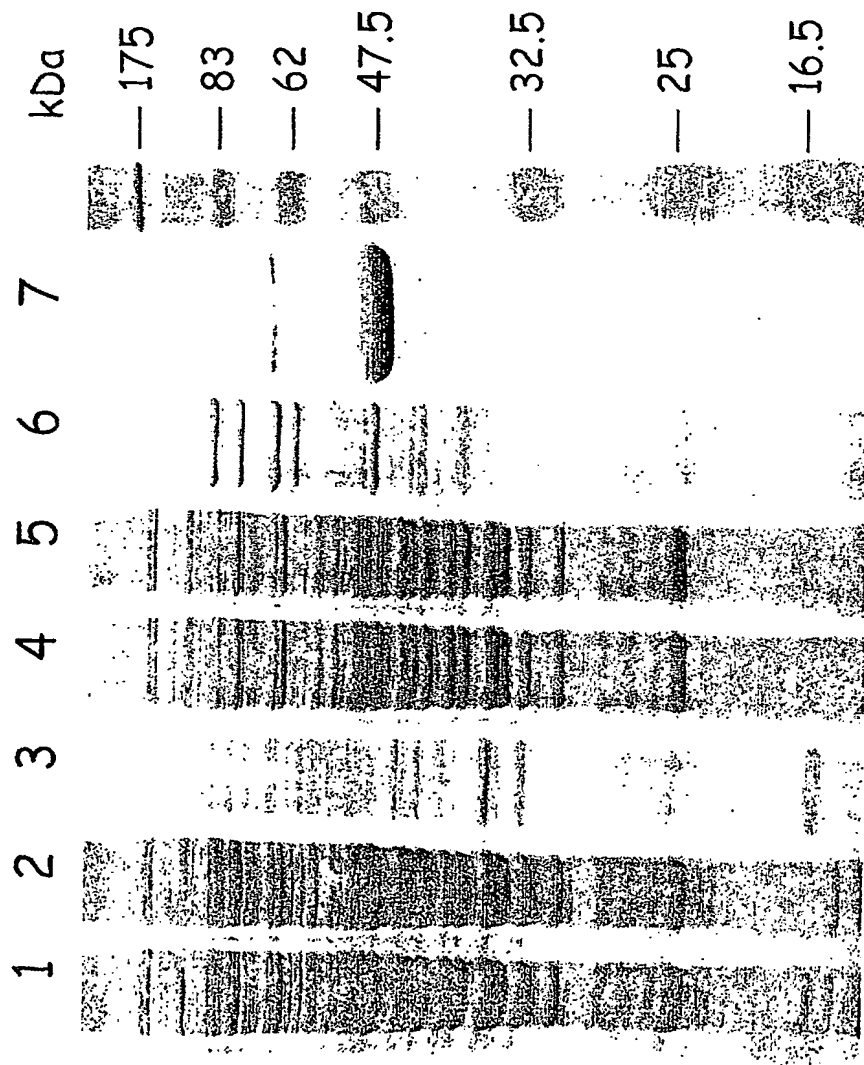


Figure 5

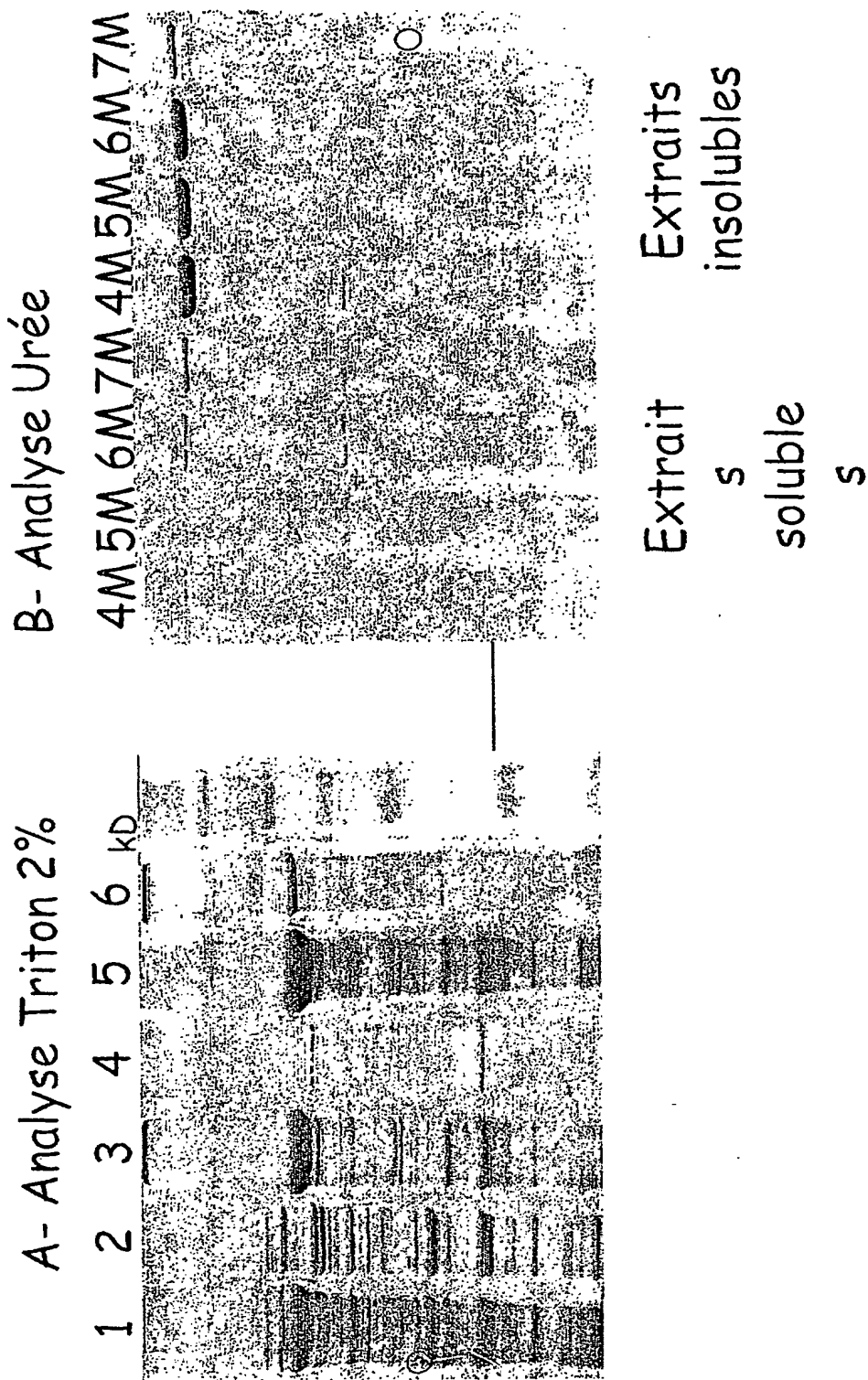


Figure 6

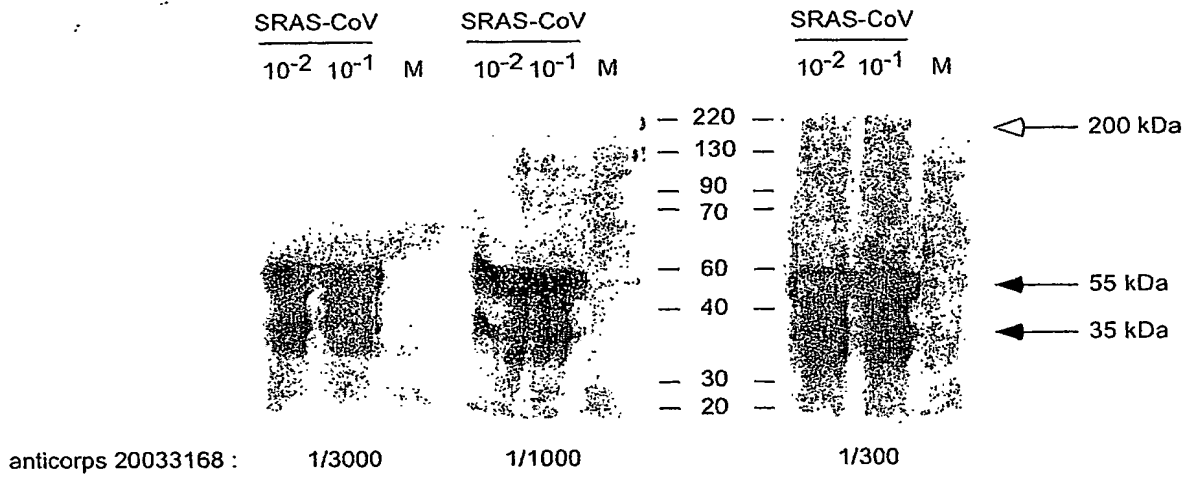


FIGURE 7

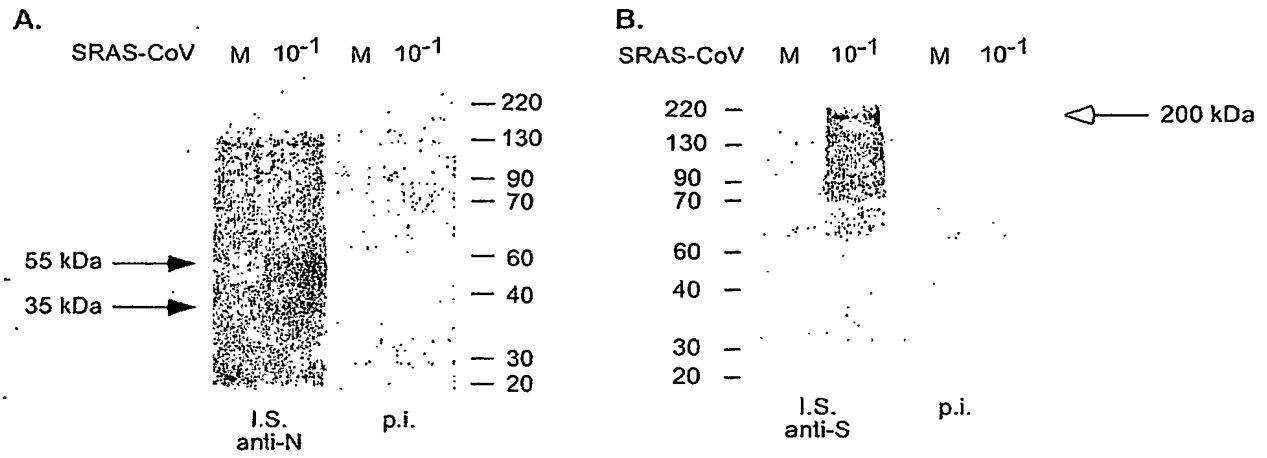
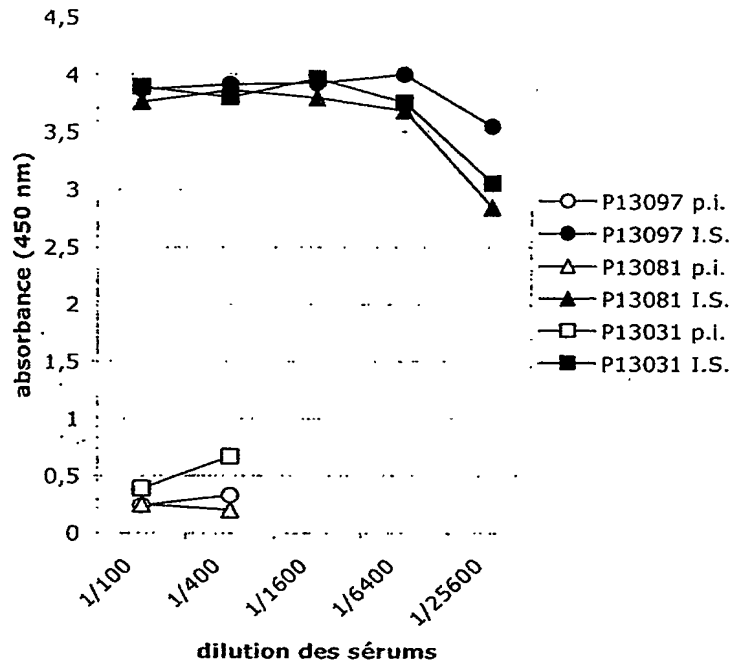


FIGURE 8

A



B

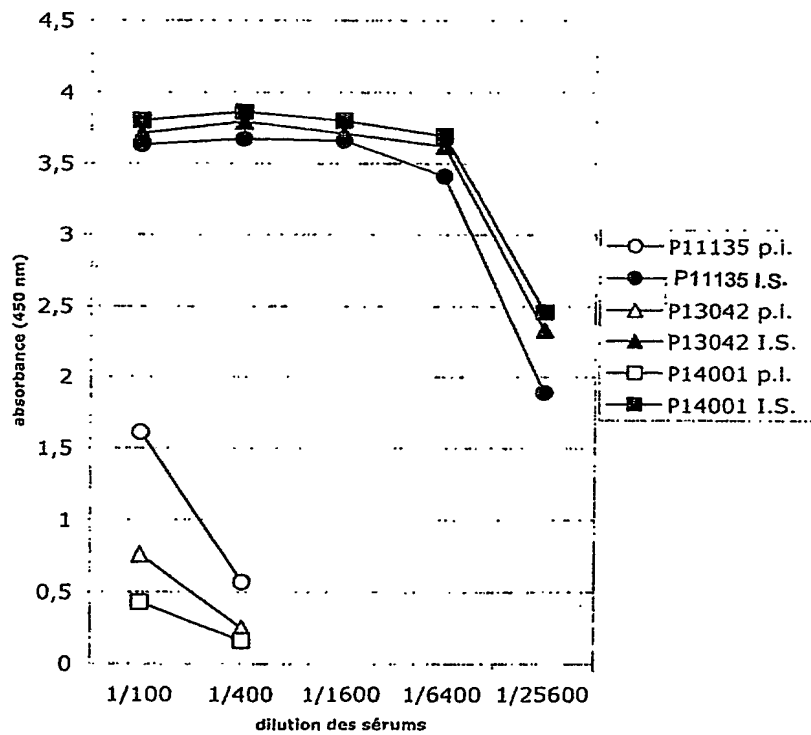
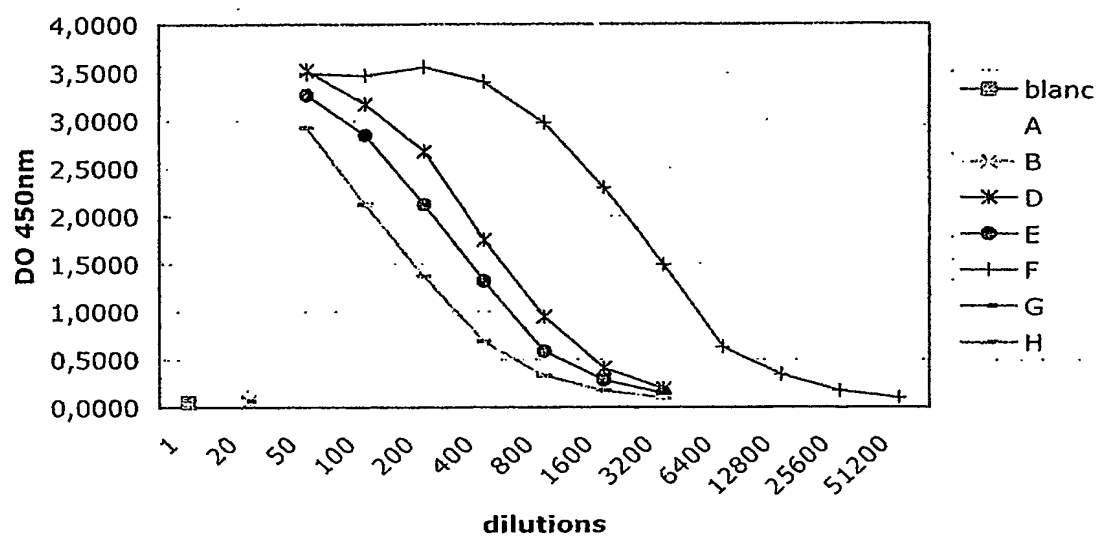


Figure 9

10/83

### ELISA-N 4 $\mu$ g



### ELISA-N 2 $\mu$ g

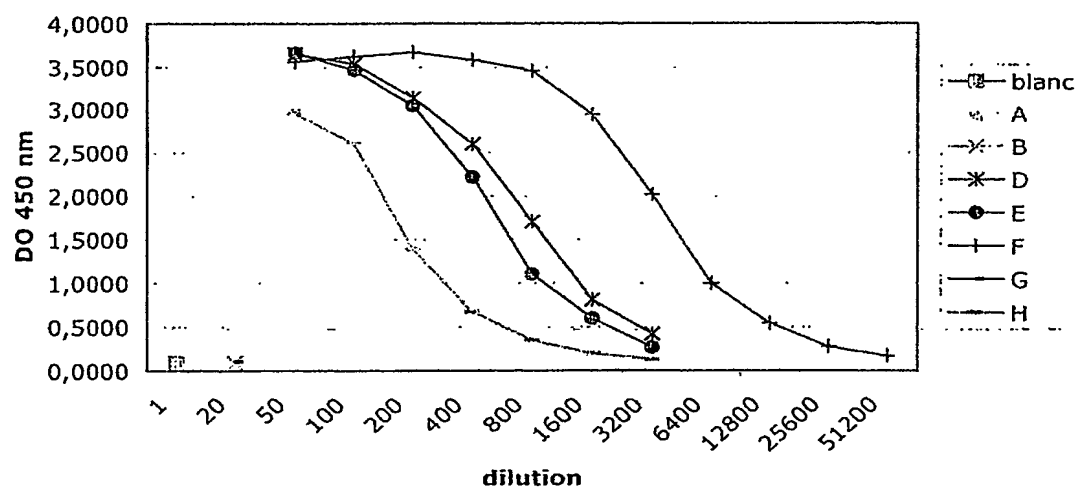


Figure 10a

11/83

# ELISA-N 1 µg

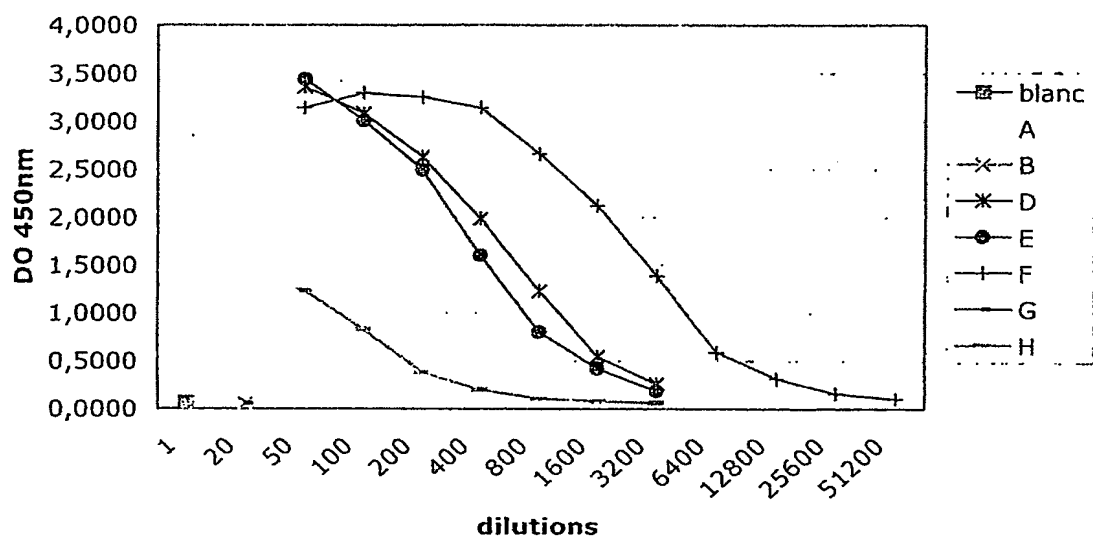


Figure 10b

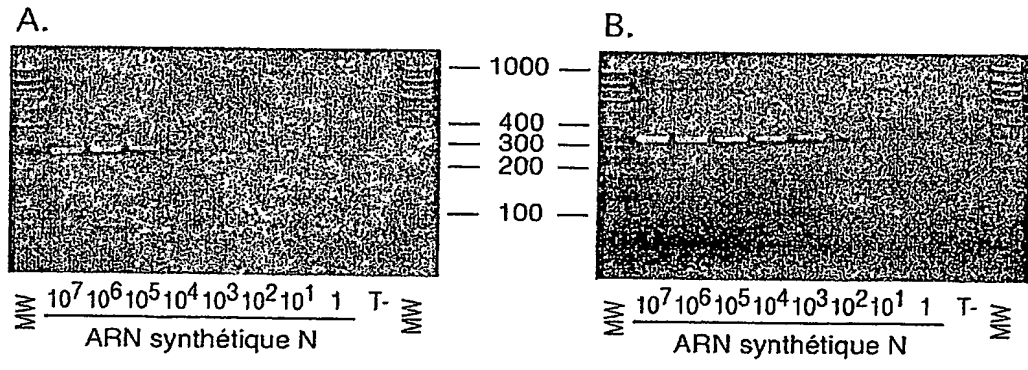


Figure 11



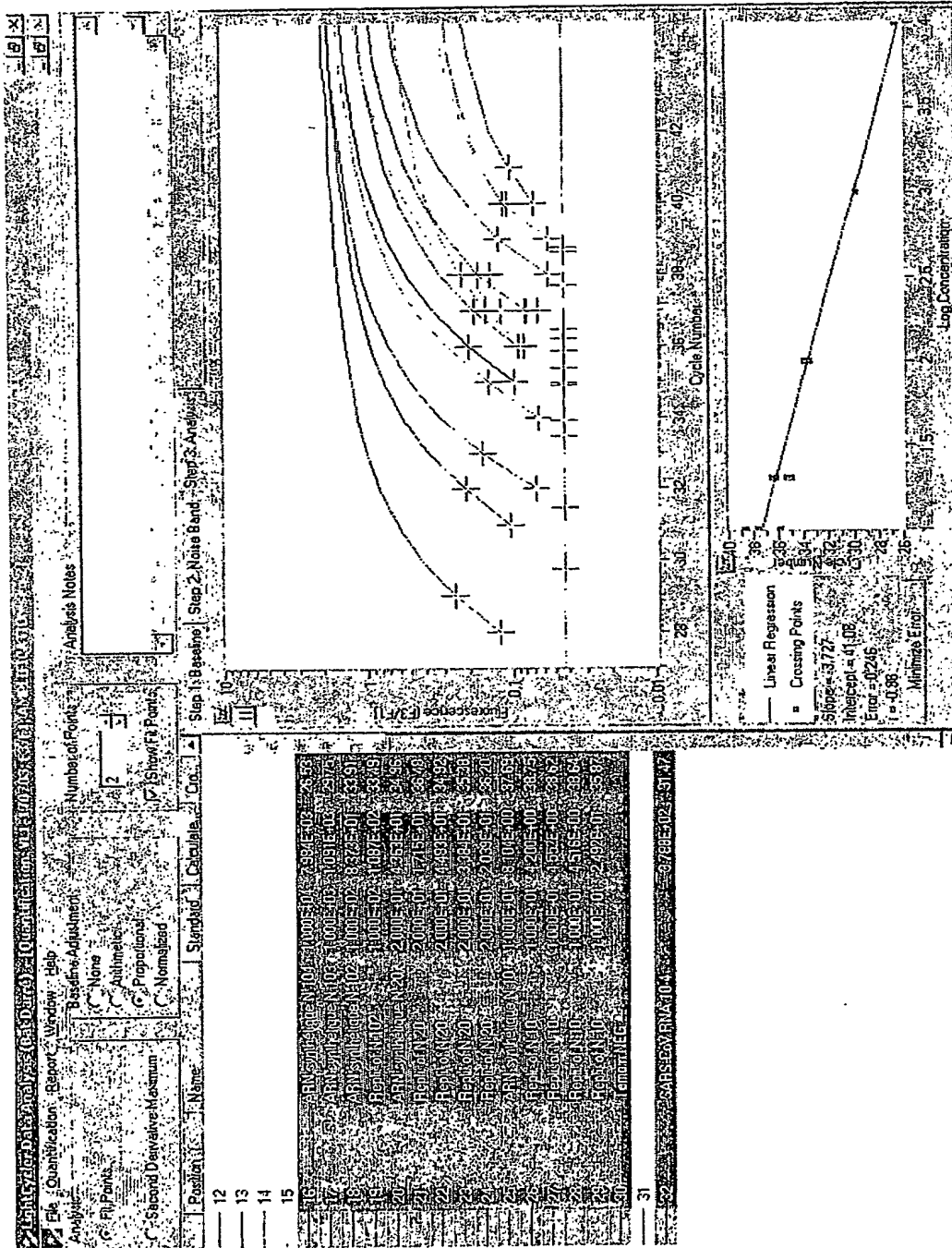


Figure 12



```

>< Sau96I
  >< PssI
    >< Pali
      >< NspIV
        >< MnlI
          >< HaeIII
            >< EcoO109I
              >< DraII>< MboII >< PmlI
                >< MnlI >< Cfr13I >< PmaCI
                  >< Ksp632I >< BsuRI > < MaeII
                    >< HinfI >< BsiZI>< EcoNI >< Eco72I
                      >< EarI >< BshI >< BslI >< BsaAI
                        >< PleI >< Eam1104I>< AsuI >< BsiYI>< BbrPI >< MnlI
TGGCTTCGGG GACTCTGTGG AAGAGGCCCT ATCGGAGGCA CGTGAACACC TCAAAAATGG CACTTGTGGT
  360          370          380          390          400          410          420

                                >< Tru9I
                                >< SfaNI
                                >< MseI
>< RmaI >< Csp6I >< BspWI >< MaeII
>< MaeI >< AluI >< AfaI >< AluI >< MaeII
CTAGTAGAGC TGGAAAAAGG CGTACTGCCC CAGCTTGAAC AGCCCTATGT GTTCATTAAA CGTTCTGATG
  430          440          450          460          470          480          490

    >< Pali
      >< HaeIII
        >< Tru9I >< GdiII >< RsaI
          >< MseI >< EaeI >< McrI ><
            >< Esp4I >< BsuRI >< Csp6I
              >< AflII >< BshI >< BsmI BsiEI ><
                CCTTAAGCAC CAATCACGGC CACAAGGTCG TTGAGCTGGT TGCAGAAATG GACGGCATTC AGTACGGTCG
                  500          510          520          530          540          550          560

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< BslI
                                >< BsiYI
                                >< MboII
                                >< MboII
>< AciI >< AfaI >< AflIII >< MunI >< AciI
TAGCGGTATA AACTGGGAG TACTCGTGCC ACATGTGGGC GAAACCCCAA TTGCATACCG CAATGTTCTT
  570          580          590          600          610          620          630

                                >< TthHB8I
                                >< TaqI
                                >< Sau3AI
                                >< NdeII
                                >< MboI
                                >< DpnII
                                >< DpnI
                                >< ClaI
                                >< Bsu15I
                                >< BspDI
                                >< BspAI
                                >< Bsp143I
                                >< Bsp106I
                                >< BsiXI >< MaeIII >
                                >< BscI>< SfaNI DdeI ><
                                >< BfrI ><
                                >< BscBI >< AluI >< BanIII
CTTCGTAAGA ACGGTAATAA GGGAGCCGGT GGTCATAGCT ATGGCATCGA TCTAAAGTCT TATGACTTAG
  640          650          660          670          680          690          700

```

FIGURE 13.2

FIGURE 13.3

17/83

```

>< MvaI      >< Hin6I      >< SduI      >< Csp45I
>< Ecl136I   >< HhaI      >< NspII     >< BstBI
>< BstOI     >< HaeII     >< HgiAI     >< Bsp119I
>< BstNI     >< Eco47III   >< Bsp1286I  >< BsiCI
>< BsiLI     >< CfoI      >< BmyI      >< Bpu14I
>< ApyI >< DdeI >< Bsp143II >< AluI   >< Alw21I   >< AsuII
CTGGTTCAC   GAGCGCTCTG ATAAGAGCTA CGAGCACCAG ACACCCTTCG AAATTAAGAG TGCCAAGAAA
      990      1000      1010      1020      1030      1040      1050

                                >< Tru9I
                                >< BsmI
                                >< MseI
                                >< BscCI
                                >< MnlI
TTTGACACTT  TCAAAGGGGA ATGCCCAAAG TTTGTGTTTC CTCTTAACTC AAAAGTCAAA GTCATTCAAC
      1060      1070      1080      1090      1100      1110      1120

>< PmlI
>< PmaCI
>< MaeII
>< Eco72I
>< BsaAI
>< BbrPI
>< AflIII   >< MnlI>< DdeI
CACGTGTTGA  AAAGAAAAAG ACTGAGGGTT TCATGGGGCG TATACGCTCT GTGTACCCTG TTGCATCTCC
      1130      1140      1150      1160      1170      1180      1190

>< SfaNI
>< MaeIII   >< AccI
ACAGGAGTGT  AACAATATGC ACTTGCTAC  CTTGATGAAA TGTAATCATT GCGATGAAGT TTCATGGCAG
      1200      1210      1220      1230      1240      1250      1260

                                >< SinI
                                >< Sau96I
                                PssI ><
                                >< Psp5II
                                >< PpuMI
                                >< NspIV
                                >< NspHII
                                >< Eco47I
                                >< DraII
                                >< Cfr13I
                                >< BsiZI
                                >< Bme18I
                                >< AvaII
                                >< AsuI
                                EcoO109I ><AflIII >
ACGTGCGACT  TTCTGAAAGC CACTTGTGAA CATTGTGGCA CTGAAAATTT AGTTATTGAA GGACCTACTA
      1270      1280      1290      1300      1310      1320      1330

                                Van91I ><
                                SinI ><
                                Sau96I ><
                                PflMI ><
                                NspIV ><
                                NspHII >
                                Eco47I ><
                                Cfr13I ><
                                BslI ><
                                BsiZI ><
                                BsiYI ><
                                Bme18I ><
                                AvaII ><
                                AsuI ><

>< MaeII
ACGTGCGACT  TTCTGAAAGC CACTTGTGAA CATTGTGGCA CTGAAAATTT AGTTATTGAA GGACCTACTA
      1270      1280      1290      1300      1310      1320      1330

                                Van91I ><
                                SinI ><
                                Sau96I ><
                                PflMI ><
                                NspIV ><
                                NspHII >
                                Eco47I ><
                                Cfr13I ><
                                BslI ><
                                BsiZI ><
                                BsiYI ><
                                Bme18I ><
                                AvaII ><
                                AsuI ><

>< RsaI
>< NspI
>< NlaIV
>< NlaIII
>< NspHI>< KpnI
>< Eco64I
>< Csp6I
>< BscBI
>< BanI
>< Asp718
>< AfaI
>< AccBI

```

FIGURE 13. 4

```

      >< Acc65I      >< SfcI      >< NlaIII      AccB7I ><
CATGTGGGTA CCTACCTACT AATGCTGTAG TGAAAATGCC ATGTCCTGCC TGTCAAGACC CAGAGATTGG
      1340      1350      1360      1370      1380      1390      1400

                                >< TthHB8I
                                >< TaqI>< MnlI
                                >< HinfI
      >< DdeI
ACCTGAGCAT AGTGTTCAG ATTATCACAA CCACTCAAAC ATTGAAACTC GACTCCGCAA GGGAGGTAGG
      1410      1420      1430      1440      1450      1460      1470

                                >< PleI      >< AciI
      >< RmaI
                                NlaIV ><
      >< MnlI
                                >< BsrI
      >< MaeI
                                >< BbvI      >< Fnu4HI      BscBI ><
ACTAGATGTT TTGGAGGCTG TGTGTTTGCC TATGTTGGCT GCTATAATAA GCGTGCCTAC TGGGTTCTCTC
      1480      1490      1500      1510      1520      1530      1540

                                XhoII ><
                                Sau3AI ><
                                NdeII ><
                                MflI ><
                                MboI ><
                                >< MaeIII
                                >< Eco31I      DpnII ><
      >< RmaI
                                >< PalI
      >< MnlI      >< HaeIII
                                >< BsrI      >< MnlI DpnI >
      >< MaeI      >< BsuRI      >< BsmAI      BstYI ><
      >< DdeI      >< BspWI      >< BsaI>< HphI      BspAI ><
      >< BshI>< BglI      >< Alw26I      Bsp143I >
GTGCTAGTGC TGATATTGGC TCAGGCCATA CTGGCATTAC TGGTGACAAT GTGGAGACCT TGAATGAGGA
      1550      1560      1570      1580      1590      1600      1610

      >< Tru9I
      >< MseI
      >< MaeII      >< Tru9I
      >< HpaI
      >< HindII
      >< HinfI >< PleI >< HincII
      >< AlwI >< DdeI      >< AflIII      >< MseI
TCTCCTTGAG ATACTGAGTC GTGAACGTGT TAACATTAAAC ATTGTTGGCG ATTTTCATTT GAATGAAGAG
      1620      1630      1640      1650      1660      1670      1680

      >< MboII
      >< BstXI      >< SfaNI
      >< HinfI
GTTGCCATCA TTTTGGCATC TTTCTCTGCT TCTACAAGTG CCTTTATTGA CACTATAAAG AGTCTTGATT
      1690      1700      1710      1720      1730      1740      1750

                                >< StyI
                                >< MaeIII
                                >< EcoT14I
                                >< Eco130I
      >< PleI
      >< MaeIII
      >< BssT1I      BslI ><
      >< HinfI>< AciI      >< BsaJI      BsiYI ><
ACAAGTCTTT CAAAACCATT GTTGAGTCCT GCGGTAAC TAAGTTACC AAGGGAAAGC CCGTAAAGG
      1760      1770      1780      1790      1800      1810      1820

      >< Sau3AI
      >< NdeII
      >< MboI
      >< DpnII
      >< DpnI >< Tru9I
      >< BspAI >< MseI
      >< Bsp143I
      >< Van91I
      >< PflMI
      >< DraIII
      >< BslI
      >< BsiYI
      >< BbvI
      >< AccB7I      Fnu4HI ><
      >< MnlI

```

FIGURE 13.5

19/83

```

TGCTTGAAC ATTGGACAAC AGAGATCAGT TTTAACACCA CTGTGTGGTT TTCCCTCACA GGCTGCTGGT
1830          1840          1850          1860          1870          1880          1890

                >< ThaI
                >< SfaNI
                >< MvnI
                >< HinPII
                >< HinPII
                >< Hin6I
                >< Hin6I
                >< HhaI
                >< HhaI
>< Sau3AI      >< HhaI
>< NdeII        >< CfoI
>< MboI         >< CfoI
>< DpnII        >< BstUI
                >< DpnI  >< BssHII
>< BspAI        >< Bsp50I
                >< Bsp143I >< AccII
GTTATCAGAT CAATTTTTCG GCGCACACTT GATGCAGCAA ACCACTCAAT TCCTGATTTG CAAAGAGCAG
1900          1910          1920          1930          1940          1950          1960

                >< TthHB8I
                >< StyI
                >< NcoI
                >< HindII
                >< HincII
                >< HinII
                >< EcoT14I
                >< Eco57I
                >< TaqI>< Eco130I
>< SalI >< DsaI
>< RtrI >< BssTII
                >< BsaHI
                >< BbiIII>< NlaIII
                >< AcyI >< HgaI
>< MaeIII
                >< BbvI
                >< MaeII >< AccI>< BsaJI
CTGTCACCAT ACTTGATGGT ATTTCTGAAC AGTCATTACG TCTTGTCGAC GCCATGGTTT ATACTTCAGA
1970          1980          1990          2000          2010          2020          2030

                >< RsaI
                >< Csp6I
                >< AfaI
                >< DdeI
>< BspMI
CCTGCTCACC AACAGTGTCA TTATTATGGC ATATGTAAC TGGTGGTCTTG TACAACAGAC TTCTCAGTGG
2040          2050          2060          2070          2080          2090          2100

                >< StuI
                >< Pali
                >< HaeIII
                >< Eco147I
                >< SduI
                >< DdeI
                >< NspII
                >< BsuRI
                >< Bsp1286I
                >< BshI
                >< BmyI
                >< AatI
                >< MnlI
                >< DdeI ><
                >< BfrI ><
TTGTCTAATC TTTTGGGCAC TACTGTTGAA AAACCTCAGGC CTATCTTTGA ATGGATTGAG GCGAAACTTA
2110          2120          2130          2140          2150          2160          2170

                >< TfiI
                >< HinfI
                >< FokI
                >< SfaNI >< BsgI
                >< Tth111I ><
GTGCAGGAGT TGAATTTCTC AAGGATGCTT GGGAGATTCT CAAATTTCTC ATTACAGGTG TTTTTCACAT
2180          2190          2200          2210          2220          2230          2240

```

FIGURE 13.6

20/83

```

Tru9I ><
MseI ><
HpaI >
HindII >
HincII >
>< Eco57I
CGTCAAGGGT CAAATACAGG TTGCTTCAGA TAACATCAAG GATTGTGTAA AATGCTTCAT TGATGTTGTT
2250.      2260      2270      2280      2290      2300      2310

>< Sau3AI
>< NdeII
>< MboI
> < MaeIII
>< FbaI
>< DpnII
>< DpnI
>< BspAI
>< Bsp143I
>< TthHB8I
>< TaqI
AACAAGGCAC TCGAAATGTG CATTGATCAA GTCACATATCG CTGGCGCAAA GTTGCGATCA CTCAACTTAG
2320      2330      2340      2350      2360      2370      2380

>< PvuII
>< MaeII
>< Bst1107I
>< BsaAI
>< BbvI
>< HphI
>< DrdI
>< AccI
>< Psp5I
>< Bsp143I
>< HhaI
>< CfoI
>< MboIBfrI
>< BspAI
>< BbsI
>< PvuII
>< Psp5I
>< NspBII
>< Fnu4HI
>< Fnu4HI
>< AluI
GTGAAGTCTT CATCGCTCAA AGCAAGGGAC TTTACCGTCA GTGTATACGT GGCAAGGAGC AGCTGCAACT
2390      2400      2410      2420      2430      2440      2450

>< Tru9I
>< NlaIV
>< MseI
>< MnlI
>< Esp4I
>< Eco64I
>< BscBI
>< NlaIII >< BanI
>< AflII
>< BbvI
>< AccBII
>< MaeIII
>< TfiI
>< HinfI
>< HphI
>< AfaI
>< ScaI
>< RsaI
>< NlaIIIMnlI ><
MnlI ><
>< Csp6I
ACTCATGCCT CTTAAGGCAC CAAAAGAAGT AACCTTTCTT GAAGGTGATT CACATGACAC AGTACTTACC
2460      2470      2480      2490      2500      2510      2520

> < XhoI
>< TthHB8I
>< TthHB8I >< TaqI
> < SlaI
> < PaeR7I
> < NspIII
>< HphI >< HinII
> < Eco88I
> < CcrI
>< Esp3I >< BsaHI
> < BcoI
>< BsmAI >< BbiII
> < AvaI >< HgaI
>< TaqI > < Ama87I >< BsmBI
>< Alw26I >< AcyI >< AluI
>< DdeI >< MnlI
TCTGAGGAGG TTGTTCTCAA GAACGGTGAA CTCGAAGCAC TCGAGACGCC CGTTGATAGC TTCACAAATG
2530      2540      2550      2560      2570      2580      2590

```

FIGURE 13.7



21/83

```

                                >< PstI >< NlaIII
                                >< HaeIII >< MnlI
                                >< BsuRI >< DdeI >< Tru9I
                                >< BshI >< BfrI >< MseI
>< AluI >< BsrI
GAGCTATCGT TGGCACACCA GTCTGTGTAA ATGGCCTCAT GCTCTTAGAG ATTAAGGACA AAGAACAATA
2600 2610 2620 2630 2640 2650 2660

                                >< VneI
                                Tru9I ><
                                >< SnoI
                                >< SduI
                                >< NspII
                                MseI ><
                                >< HgiAI
                                Bsp1286I >< BslI ><
                                >< BsiYI ><
                                >< BmyI
                                >< ApaLI
                                >< Tru9I >< Alw44I
                                >< MseI >< Alw21I
CTGCGCATTG TCTCCTGGTT TACTGGCTAC AAACAATGTC TTTCGCTTAA AAGGGGGTGC ACCAATTAAA
2670 2680 2690 2700 2710 2720 2730

                                >< TfiI
>< MaeIII >< MboII >< MaeIII >< HinfI AluI ><
GGTGTAACTT TTGGAGAAGA TACTGTTTGG GAAGTTCAAG GTTACAAGAA TGTGAGAATC ACATTTGAGC
2740 2750 2760 2770 2780 2790 2800

                                >< RsaI
                                >< NlaIV
                                MaeIII ><
                                >< MspI >< KpnI
                                >< HpaII
                                >< HapII
                                >< Eco64I
                                >< Csp6I
                                >< SduI
                                >< NspII >< TfiI >< BscBI
                                >< HgiAI >< BanI
                                >< Bsp1286I >< Asp718
                                >< BmyI >< HinfI >< AfaI
                                >< Alw21I >< AccB1I
                                >< AccI >< Acc65I
TTGATGAACG TGTGACAAA GTGCTTAATG AAAAGTGCTC TGTCTACACT GTTGAATCCG GTACCGAAGT
2810 2820 2830 2840 2850 2860 2870

                                >< Sau3AI
                                >< NdeII
                                >< MboI
                                >< DpnII
                                >< DpnI
                                >< NspI
                                >< NspHI
                                >< MboII >< BspAI
                                >< NlaIII >< BsrI >< Bsp143I
                                >< DdeI >< MnlI >< AlwNI >< BbsI >< AlwNI
TACTGAGTTT GCATGTGTTG TAGCAGAGGC TGTGTGAAG ACTTTACAAC CAGTTTCTGA TCTCCTTACC
2880 2890 2900 2910 2920 2930 2940

                                >< Sau3AI
                                >< NdeII
                                >< MboI
                                >< DpnII
                                >< DpnI
                                >< BspAI

```

FIGURE 13.8

22/83

```

>< NlaIII>< Bsp143I          >< AluI          >< SfaNI
AACATGGGTA TTGATCTTGA TGAGTGGAGT GTAGCTACAT TCTACTTATT TGATGATGCT GGTGAAGAAA
  2950      2960      2970      2980      2990      3000      3010

                                >< SfaNI
                                >< MnlI
                                >< Ksp632I          >< MnlI
                                >< EarI          >< MboII
>< MboII          >< GsuI          >< MnlI          >< Eam1104I          >< MboII
    >< BsaAI
>< HphI >< MaeII>< BpmI          >< MnlI          >< Eam1104I          >< MboII
ACTTTTCATC ACGTATGTAT TGTTCTTTT ACCCTCCAGA TGAGGAAGAA GAGGACGATG CAGAGTGTGA
  3020      3030      3040      3050      3060      3070      3080

                                >< RsaI
                                >< RsaI
                                >< NlaIII
                                >< MnlI          >< FokI
                                >< Csp6I          Eco31I ><
                                >< Csp6I          >< MamI BsmAI ><
                                >< AfaI          >< BsiBI BsaI ><
                                >< AfaI          >< BsaB1Alw26I ><
>< MboII          >< MboII          >< AfaI          >< BsaB1Alw26I ><
GGAAGAAGAA ATTGATGAAA CCTGTGAACA TGAGTACGGT ACAGAGGATG ATTATCAAGG TCTCCCTCTG
  3090      3100      3110      3120      3130      3140      3150

>< NlaIV>< PvuII>< XmnI
>< Eco64I >< Psp5I          >< TthHB8I
>< MnlI >< DdeI          >< TaqI          >< MnlI          >< MboII
>< BscBI>< NspBII >< MnlI          >< Ksp632I          >< MboII >< MboII
>< BanI          >< MnlI          >< EarI          >< BsrI
>< AccBII >< AluI >< Asp700I          >< Eam1104I >< MboII>< BbsI
GAATTTGGTG CCTCAGCTGA AACAGTTCGA GTTGAGGAAG AAGAAGAGGA AGACTGGCTG GATGATACTA
  3160      3170      3180      3190      3200      3210      3220

                                >< Tru9I
                                >< MseI          >< Eco57I
>< FokI          >< BsrI>< MboII BsrI ><
>< DdeI          CTGAGCAATC AGAGATTGAG CCAGAACCCAG AACCTACACC TGAAGAACCA GTTAATCAGT TTACTGGTTA
  3230      3240      3250      3260      3270      3280      3290

                                >< MnlI
>< Tru9I          >< Tru9I          >< HindII>< Tru9I          >< DraIII
>< MseI          >< MseI          >< HincII>< MseI          >< BspWI
>< DraI          TTTAAACTT ACTGACAATG TTGCCATTAA ATGTGTTGAC ATCGTTAAGG AGGCACAAAG TGCTAATCCT
  3300      3310      3320      3330      3340      3350      3360

                                >< VneI
                                >< SnoI
                                >< SduI
                                >< NspII
                                >< HgiAI
                                >< Bsp1286I
                                >< BmyI
                                >< ApaLI
                                >< Alw44I
>< BbvI          >< HphI          >< NlaIII          >< Alw21I
    >< Fnu4HI          >< BspMI
ATGGTGATTG TAAATGCTGC TAACATACAC CTGAAACATG GTGGTGGTGT AGCAGGTGCA CTCAACAAGG
  3370      3380      3390      3400      3410      3420      3430

                                >< Sau96I
                                >< Pali
                                >< NspIV
                                >< HaeIII
                                >< Cfr13I
>< NlaIV

```

FIGURE 13.9

23/83

```

>< Eco64I
>< BscBI
>< BanI
>< AccB1I>< NlaIII
CAACCAATGG TGCCATGCAA AAGGAGAGTG ATGATTACAT TAAGCTAAAT GGCCCTCTTA CAGTAGGAGG
3440 3450 3460 3470 3480 3490 3500

>< BsuRI
>< Tru9I
>< MseI
>< AluI
>< AsuI
>< MnlI
>< SinI
>< Sau96I
>< NspIV
>< NspHI>< NspHII
>< Eco47I
>< Cfr13I
>< NlaIII
>< BspMI
>< BsiZI
>< Bme18I
>< AvaII MnlI ><
>< DdeI
>< NspI>< AsuI FokI ><
GTCTTGTTTG CTTTCTGGAC ATAATCTTGC TAAGAAGTGT CTGCATGTTG TTGGACCTAA CCTAAATGCA
3510 3520 3530 3540 3550 3560 3570

>< Tru9I
>< HphI> < MseI
>< Esp4I
>< AluI
>< NdeI
>< AflIII>< Fnu4HI
>< BbvI
GGTGAGGACA TCCAGCTTCT TAAGGCAGCA TATGAAAATT TCAATTCACA GGACATCTTA CTTGCACCAT
3580 3590 3600 3610 3620 3630 3640

RsaI ><
Csp6I ><
AfaI ><
>< Eco57I
>< BcgI
TGTTGTCAGC AGGCATATTT GGTGCTAAAC CACTTCAGTC TTTACAAGTG TCGGTGCAGA CGGTTCGTAC
3650 3660 3670 3680 3690 3700 3710

>< BsgI
>< BcgI/a
>< BspMI
>< AluI
>< NlaIII
ACAGGTTTAT ATTGCAGTCA ATGACAAAGC TCTTTATGAG CAGGTTGTCA TGGATTATCT TGATAACCTG
3720 3730 3740 3750 3760 3770 3780

>< RmaI
>< MaeI
>< MnlI
>< Eco57I
>< NlaIV
>< BscBI
>< TfiI
>< MboII
AAGCCTAGAG TGGAAGCACC TAAACAAGAG GAGCCACCAA ACACAGAAGA TTCCAAAACCT GAGGAGAAAT
3790 3800 3810 3820 3830 3840 3850

>< Tru9I
>< StuI
>< Pali
>< MseI
>< MnlI
>< MaeIII
>< HaeIII
>< Eco065I
>< Eco147I
>< Eco91I
>< BsuRI
>< BstXI ><
>< BshI
>< BstPI
>< AatI
>< BstEII
CTGTCGTACA GAAGCCTGTC GATGTGAAGC CAAAAATTAA GGCCTGCATT GATGAGGTTA CCACAACACT
3860 3870 3880 3890 3900 3910 3920

TfiI ><
NlaIII ><
HinfI ><
>< DdeI
>< EcoRV
>< HindIII

```

FIGURE 13 10

```

    >< BsrI      >< MboII      >< MaeIII      >< Eco32I      >< AluI
GGAAGAAACT AAGTTTCTTA CCAATAAGTT ACTCTTGTTT GCTGATATCA ATGGTAAGCT TTACCATGAT
    3930          3940          3950          3960          3970          3980          3990

    >< NspI
    >< NspHI
    >< NlaIII
    >< MnlI      >< SfaNI
    >< DdeI      >< EcoNI
    >< MboII >< BslI      >< NlaIII
    >< DdeI      >< BfrI      >< HphI      >< BsiYI      >< FokI
TCTCAGAACA TGCTTAGAGG TGAAGATATG TCTTTCCTTG AGAAGGATGC ACCTTACATG GTAGGTGATG
    4000          4010          4020          4030          4040          4050          4060

    >< SpeI
    >< RmaI
    >< MaeI      >< EcoRV>< HphI      >< SfaNI
    >< HphI      >< Eco32I      >< MnlI      >< DdeI
TTATCACTAG TGGTGATATC ACTTGTGTTG TAATACCCTC CAAAAAGGCT GGTGGCACTA CTGAGATGCT
    4070          4080          4090          4100          4110          4120          4130

    >< ScrFI
    >< RsaI
    >< MvaI
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< Csp6I >< EcoNI
    >< BstOI
    >< BstNI
    >< BsiLI
    >< BsaJI
    >< BsaAI      >< BslI
    >< MaeII>< ApyI
    >< AfaI      >< BsiYI
    >< AluI      >< MboII      >< BsrI
CTCAAGAGCT TTGAAGAAAG TGCCAGTTGA TGAGTATATA ACCACGTACC CTGGACAAGG ATGTGCTGGT
    4140          4150          4160          4170          4180          4190          4200

    >< Tru9I
    >< MseI
    >< DdeI      >< Esp4I
    >< MnlI      >< BspWI
    >< FokI      >< AluI      >< AflII      >< Eco57I >< AfaI
TATACACTTG AGGAAGCTAA GACTGCTCTT AAGAAATGCA AATCTGCATT TTATGTACTA CCTTCAGAAG
    4210          4220          4230          4240          4250          4260          4270

    >< ScrFI
    >< MvaI
    >< EcoRII
    >< Ecl136I
    >< XmnI      >< RmaI      >< DsaV      NlaIII ><
    >< Ksp632I      >< TfiI>< MboII      >< BstOI      Ksp632I ><
    >< EarI      >< Eam1104I      >< MaeI      >< BstNI      >< EarI
    >< DdeI      >< HinfI      >< BsiLI      Eam1104I ><
    >< BspWI      >< Asp700I      >< ApyI      BsmAI ><
    >< Alw26I ><
CACCTAATGC TAAGGAAGAG ATTCTAGGAA CTGTATCCTG GAATTTGAGA GAAATGCTTG CTCATGCTGA
    4280          4290          4300          4310          4320          4330          4340

    >< VspI      >< Zsp2I
    >< Tru9I      >< Ppu10I
    >< MseI      >< NsiI
    >< MboII      >< NlaIII      >< FokI
    >< Eco57I      >< Mph1103I      >< FokI

```

FIGURE 13. 11

```

                >< AsnI          >< EcoT22I          >< BspWI
                >< AseI          >< AvaIII          >< BglI          >< MaeII
AGAGACAAGA AAATTAATGC CTATATGCAT GGATGTTAGA GCCATAATGG CAACCATCCA ACGTAAGTAT
    4350          4360          4370          4380          4390          4400          4410

                >< SfaNI
                >< Tru9I          > < HindII          >< TfiI          >< SpeI
                >< MseI          > < HincII>< MboII          >< RmaI
                >< MnlI          >< DrdI >< HinfI          >< MaeI
AAAGGAATTA AAATTCAAGA GGGCATCGTT GACTATGGTG TCCGATTCTT CTTTATACT AGTAAAGAGC
    4420          4430          4440          4450          4460          4470          4480

                >< MaeIII
>< SfcI          >< Fnu4HI          >< MnlI
>< AluI          >< AluI          >< AciI          MaeIII ><
CTGTAGCTTC TATTATTACG AAGCTGAACT CTCTAAATGA GCCGCTTGTC ACAATGCCAA TTGGTTATGT
    4490          4500          4510          4520          4530          4540          4550

                >< Thai
                >< MvnI
                >< MboII
                >< HinPII
>< HinPII
>< Hin6I
>< Hin6I
>< HhaI
>< HhaI
>< Tru9I
>< NlaIII          >< Fnu4HI
>< MseI          >< CfoI
>< MnlI          >< CfoI
>< Ksp632I          >< BstUI
>< EarI          >< BssHII>< BspWI          >< Tru9I
>< Eam1104I          >< Bsp50I          >< MseI
>< BbvI          >< AccII          >< AluI          HphI ><
GACACATGGT TTTAATCTTG AAGAGGCTGC GCGCTGTATG CGTTCTCTTA AAGCTCCTGC CGTAGTGTCA
    4560          4570          4580          4590          4600          4610          4620

                >< MaeIII
>< SfaNI          >< AlwNI          >< MnlI >< MnlI>< DdeI
GTATCATCAC CAGATGCTGT TACTACATAT AATGGATACC TCACTTCGTC ATCAAAGACA TCTGAGGAGC
    4630          4640          4650          4660          4670          4680          4690

                >< SinI
                >< Sau96I
                >< NspIV
                >< NspHII
>< SduI          >< Eco47I
>< NspII          >< Cfr13I
>< HgiAI          >< Bsi2I
>< Bsp1286I          >< Bme18I          >< RsaI
>< BmyI          >< AvaII          >< Csp6I
>< Alw21I          >< AsuI          >< AfaI
ACTTTGTAGA AACAGTTTCT TTGGCTGGCT CTTACAGAGA TTGGTCCTAT TCAGGACAGC GTACAGAGTT
    4700          4710          4720          4730          4740          4750          4760

                > < TthHB8I
                > < TaqI
                >< SduI
                >< NspII
                >< Van91I          >< RsaI          >< PflMI          >< Eco24I
                >< MseI          >< HphI          >< BslI          >< Bsp1286I
                >< Esp4I          >< Csp6I          >< BsiYI          >< BmyI          GsuI ><

```

FIGURE 13.12

26/83

```

      >< AflIII >< MaeIII      >< AfaI >< AccB7I >< BanIIBpmI ><
AGGTGTTGAA TTTCTTAAGC GTGGTGACAA AATTGTGTAC CACACTCTGG AGAGCCCCGT CGAGTTTCAT
  4770      4780      4790      4800      4810      4820      4830

                                >< Tru9I
                                >< PleI >< EcoNI
                                >< MnlI >< BslI
                                >< BsmAI >< BsiYI
      >< MnlI      >< HphI      >< HinfI>< Alw26I>< AciI >< MseI
CTTGACGGTG AGGTTCTTTC ACTTGACAAA CTAAAGAGTC TCTTATCCCT GCGGGAGGTT AAGACTATAA
  4840      4850      4860      4870      4880      4890      4900

                                >< AluI      >< NdeI
AAGTGTTTAC AACTGTGGAC AACACTAATC TCCACACACA GCTTGTGGAT ATGTCTATGA CATATGGACA
  4910      4920      4930      4940      4950      4960      4970

      >< SinI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< Eco47I
      >< Cfr13I
      >< BsiZI
      >< Bme18I
      >< AvaII
      >< AsuI
                                >< MaeIII >< Tru9I >< MnlI
                                >< FokI >< MseI >< BspHI
GCAGTTTGGT CCAACATACT TGGATGGTGC TGATGTTACA AAAATTAAAC CTCATGTAAA TCATGAGGGT
  4980      4990      5000      5010      5020      5030      5040

                                > < TthHB8I
                                > < TaqI
      >< RsaI
      > < RmaI      >< SnaBI      >< ScaI
      > < MaeI      >< MaeII >< HindIII >< RsaI
      >< Csp6I      >< Eco105I      >< Csp6I
      >< AfaI      >< BsaAI >< AluI >< AfaI
AAGACTTTCT TTGTACTACC TAGTGATGAC ACACTACGTA GTGAAGCTTT CGAGTACTAC CATACTCTTG
  5050      5060      5070      5080      5090      5100      5110

      >< RsaI
      >< NspI
      >< NspHI
      >< NlaIII
      > < Csp6I      >< Tru9I
      >< AflIII      >< MseI
      >< AfaI      >< DraI
ATGAGAGTTT TCTTGGTAGG TACATGTCTG CTTTAAACCA CACAAAGAAA TGGAAATTTT CTCAAGTTGG
  5120      5130      5140      5150      5160      5170      5180

      >< Tru9I >< Tru9I      >< RmaI
      >< MseI >< MseI      >< MunI >< MaeI      AluI >
TGGTTTAACT TCAATTAAAT GGGCTGATAA CAATTGTTAT TTGTCTAGTG TTTTATTAGC ACTTCAACAG
  5190      5200      5210      5220      5230      5240      5250

                                >< SfaNI
                                >< SduI
                                >< NspII
                                >< Eco24I
                                >< Bsp1286I
                                >< BmyI      HphI >
                                >< BbvI Fnu4HI ><
                                >< BanII >< BspWI
      >< MnlI

```

FIGURE 13.13

27/83

```

CTTGAAGTCA AATTCAATGC ACCAGCACTT CAAGAGGCTT ATTATAGAGC CCGTGCTGGT GATGCTGCTA
  5260          5270          5280          5290          5300          5310          5320

>< VneI
>< SnoI
    >< SduI
    >< NspII
    >< HgiAI
    >< Bsp1286I
    >< BmyI
>< ApaLI
>< Alw44I
    >< Alw21I
ACTTTTGTGC ACTCATACTC GCTTACAGTA ATAAAACTGT TGGCGAGCTT GGTGATGTCA GAGAACTAT
  5330          5340          5350          5360          5370          5380          5390
                                >< AluI
                                >< HphI
                                MboII ><

    > < SphI
    > < PaeI
    > < NspI
    > < NspHI >< TfiI
    >< SfcI > < NlaIII>< HinfI
GACCCATCTT CTACAGCATG CTAATTTGGA ATCTGCAAAG CGAGTTCTTA ATGTGGTGTG TAAACATTGT
  5400          5410          5420          5430          5440          5450          5460
                                >< Tru9I
                                >< MseI
                                >< RsaI
                                > < Csp6I      Esp4I >
                                >< AfaI      AflIII >
GGTCAGAAAA CTACTACCTT AACGGGTGTA GAAGCTGTGA TGTATATGGG TACTCTATCT TATGATAATC
  5470          5480          5490          5500          5510          5520          5530
                                >< RsaI
                                >< MboII
                                >< RmaIHinfI ><
                                >< Csp6I
>< Tru9I
>< MseI
    >< SfaNI
    >< NlaIII
TTAAGACAGG TGTTCCTATT CCATGTGTGT GTGGTCGTGA TGCTACACAA TATCTAGTAC AACAGAGTC
  5540          5550          5560          5570          5580          5590          5600
                                >< RsaI
                                >< Csp6I
                                >< AfaI
                                >< MaeI >< BbsI
                                >< AfaI
                                >< RsaI
                                >< Csp6I
                                >< AfaI
>< PleI
>< BsgI
    > < DdeI
    >< BspWI >< BspMI
TTCTTTTGTT ATGATGTCTG CACCACCTGC TGAGTATAAA TTACAGCAAG GTACATTCTT ATGTGCGAAT
  5610          5620          5630          5640          5650          5660          5670
                                >< RsaI
                                >< DdeI
                                >< BsmAI
                                >< BsaI
                                >< Alw26I
                                MnlI ><
                                HphI >
>< RsaI
> < MaeIII
>< Csp6I
>< AfaI >< BsrI
GAGTACACTG GTAACATCA GTGTGGTCAT TACTCTATA TAACTGCTAA GGAGACCOCTC TATCGTATTG
  5680          5690          5700          5710          5720          5730          5740
                                >< Eco31I
                                >< DdeI
                                >< BsmAI
                                >< BsaI
                                >< Alw26I
                                MnlI ><
                                HphI >
                                >< RsaI
                                >< MaeIII
                                >< Eco47I
                                >< Cfr13I
                                >< Bsi2I
                                >< Bme18I

    >< SstI
    >< SduI
    >< SacI
    >< NspII
    >< HgiAI
    >< Eco24I
    >< Ecl136II
    >< Bsp1286I
    >< BmyI
    >< SinI
    >< Sau96I
    >< NspIV
    >< NspHII
    > < RsaI
    >< Eco47I
    >< Cfr13I
    >< Bsi2I
    >< Bme18I

```

FIGURE 13. 14

28/83

```

    >< BanII
    >< Alw21I
    >< AluI
ACGGAGCTCA CCTTACAAAG ATGTCAGAGT ACAAAGGACC AGTGACTGAT GTTTTCTACA AGGAAACATC
    5750          5760          5770          5780          5790          5800          5810

    >< TthHB8I
    >< TaqI >< MaeIII
TTACACTACA ACCATCAAGC CTGTGTCGTA TAAACTCGAT GGAGTTACTT ACACAGAGAT TGAACCAAAA
    5820          5830          5840          5850          5860          5870          5880

    >< RsaI
    >< Csp6I
    >< SfcI >< BbvI
    >< Fnu4HI
    >< AfaI
TTGGATGGGT ATTATAAAAA GGATAATGCT TACTATACAG AGCAGCCTAT AGACCTTGTA CCAACTCAAC
    5890          5900          5910          5920          5930          5940          5950

    Tru9I ><
    SwaI ><
    MseI ><
    MamI ><
    DraI ><
    BsiBI ><
    BsaBI ><
    > < NspI
    > < NspHI
    > < NlaIII
    >< AflIII
CATTACCAAA TGCGAGTTTT GATAATTTCA AACTCACATG TTCTAACACA AAATTTGCTG ATGATTTAAA
    5960          5970          5980          5990          6000          6010          6020

    >< MboII
    >< AluI
    >< AluI>< MaeIII
TCAAATGACA GGCTTCACAA AGCCAGCTTC ACGAGAGCTA TCTGTACAT TCTTCCCAGA CTTGAATGGC
    6030          6040          6050          6060          6070          6080          6090

    >< SfcI
GATGTAGTGG CTATTGACTA TAGACACTAT TCAGCGAGTT TCAAGAAAGG TGCTAAATTA CTGCATAAGC
    6100          6110          6120          6130          6140          6150          6160

    >< Tru9I
    >< ScrFI
    >< MvaI
    >< MseI
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< BstOI
    >< BstNI
    >< BsiLI
    >< MunI
    >< BstXI
    >< ApyI
    >< MaeII
    >< DraIII
    >< BstXI
CAATTGTTTG GCACATTAAC CAGGCTACAA CCAAGACAAC GTTCAAACCA AACACTTGGT GTTTACGTTG
    6170          6180          6190          6200          6210          6220          6230

    > < RsaI
    >< Csp6I
    > < AfaI>< BsrI
    >< MboII ><
    >< BbsI
TCTTTGGAGT ACAAAGCCAG TAGATACTTC AAATTCATTT GAAGTTCTGG CAGTAGAAGA CACACAAGGA
    6240          6250          6260          6270          6280          6290          6300

    >< HindII
    >< HincII
    >< MboII
    >< MnlI
    >< Eco57I
ATGGACAATC TTGCTTGTGA AAGTCAACAA CCCACCTCTG AAGAAGTAGT GGAAAATCCT ACCATACAGA
    6310          6320          6330          6340          6350          6360          6370

```

FIGURE 13.15



29/83

```

>< MaeIII
>< MaeII
AGGAAGTCAT AGAGTGTGAC GTGAAAACCTA CCGAAGTTGT AGGCAATGTC ATACTTAAAC CATCAGATGA
6380      6390      6400      6410      6420      6430      6440

>< XhoII
>< Sau3AI
>< NlaIII
>< NdeII
>< MflI
>< MboI
>< DpnII
>< DpnI
>< BstYI
>< BspAI
>< Tru9I
>< MseI
>< BspHI >< Bsp143I>< Fnu4HI
>< MaeIII >< MnlI >< BbvI >< AlwI
AGGTGTTAAA GTAACACAAG AGTTAGGTCA TGAGGATCTT ATGGCTGCCTT ATGTGGAAAA CACAAGCATT
6450      6460      6470      6480      6490      6500      6510

>< SauI
>< RmaI
>< MstII
>< MaeI
>< Eco81I
>< DdeI
>< CvnI
>< Bsu36I
>< Bse21I
>< BfrI> < Tru9I
>< AxyI> < MseI>< MunI
>< NlaIII
>< Tru9I
>< MseI
>< AluI
>< AocI >< DraI >< BbvI Fnu4HI ><
ACCATTAAGA AACCTAATGA GCTTTCACCTA GCCTTAGGTT TAAAAACAAT TGCCACTCAT GGTATTGCTG
6520      6530      6540      6550      6560      6570      6580

>< VspI >< StyI
>< Tru9I >< EcoT14I
>< MseI >< Eco130I
>< AsnI >< BssT1I
>< AseI >< BsaJI
>< DdeI
>< BslI
>< BsiYI
>< BfrI >< Fnu4HI
CAATTAATAG TGTTCCTTGG AGTAAAATTT TGGCTTATGT CAAACCATTC TTAGGACAAG CAGCAATTAC
6590      6600      6610      6620      6630      6640      6650

>< HinPII
>< Hin6I
>< HhaI
>< DdeI
>< Tru9I
>< MaeII>< MseI
>< DraIII
>< BbvI >< CfoI
>< AflIII
AACATCAAAT TGCCTAAGA GATTAGCACA ACGTGTGTTT AACAATTATA TGCCTTATGT GTTTACATTA
6660      6670      6680      6690      6700      6710      6720

>< RsaI >< RsaI>< XbaI
>< Csp6I >< Csp6I >< RmaI
>< MunI >< AfaI >< AfaI >< MaeI >< AluI
TTGTTCCAAT TGTGTACTTT TACTAAAAGT ACCAATTCTA GAATTAGAGC TTCCTACCT ACAACTATTG
6730      6740      6750      6760      6770      6780      6790

>< VspI
>< Tru9I
>< NaeI
>< MspI
>< MseI

```

FIGURE 13. 16

30/83

```

                                >< HpaII
                                >< HapII
                                >< Cfr10I >< FokI
                                >< AsnI
                                >< AseI>< HphI>< MaeIII
                                >< Tru9I
                                >< MseI
                                >< SfaNI
CTAAAAATAG TGTTAAGAGT GTTGCTAAAT TATGTTTGGA TGCCGGCATT AATTATGTGA AGTCACCCAA
6800        6810        6820        6830        6840        6850        6860

                                >< Tru9I    >< DdeI    MaeIII >
                                >< MseI    >< BfrI    >< BbvI
ATTTTCTAAA TTGTTACAA TCGCTATGTG GCTATTGTTG TTAAGTATTT GCTTAGGTTT TCTAATCTGT
6870        6880        6890        6900        6910        6920        6930

                                >< SduI
                                >< NspII
                                >< HgiAI
                                >< Bsp1286I
                                >< BmyI
                                >< Alw21I
                                > < RsaI
                                >< Csp6I
                                >< Fnu4HI > < AfaI
GTAAGTCTG CTTTGGTGT ACTCTTATCT AATTTTGGTG CTCCTTCTTA TTGTAATGGC GTTAGAGAAT
6940        6950        6960        6970        6980        6990        7000

                                Tru9I ><
                                MseI ><
                                >< Tru9I    > < MaeIII
                                >< MseI    >< MaeII
                                >< Fnu4HI
                                >< BbvI >
TGTATCTTAA TTCGTCTAAC GTTACTACTA TGGATTCTCT TGAAGGTTCT TTTCTTGCA GCATTGTTT
7010        7020        7030        7040        7050        7060        7070

                                > < TfiI
                                >< MamI
                                >< HinfI
                                >< BsiBI
                                >< XmnI>< MaeIII
                                >< PleI>< HinfI >< BsaBI >< AluI >< Asp700I
                                >< AfaI ><
AAGTGGATTA GACTCCCTTG ATTCTTATCC AGCTCTTGAA ACCATTTCAGG TGACGATTTC ATCGTACAAG
7080        7090        7100        7110        7120        7130        7140

                                >< Pali
                                >< NspBII
                                >< HaeIII
                                >< GdiII
                                >< Fnu4HI
                                >< EaeI
                                >< DdeI
                                >< BsuRI
                                >< RmaI
                                >< BshI >< BslI
                                >< MaeI
                                >< AciI>< BsiYI
CTAGACTTGA CAATTTTAGG TCTGGCCGCT GAGTGGGTTT TGGCATATAT GTTGTTTACA AAATTCTTTT
7150        7160        7170        7180        7190        7200        7210

                                >< BspMI
                                >< AluI
                                >< RmaI
                                >< MaeI
ATTATTAGG TCTTTCAGCT ATAATGCAGG TGTCTTTGG CTATTTTGCT AGTCATTTC TCAGCAATTC
7220        7230        7240        7250        7260        7270        7280

                                RsaI ><
                                >< MboII
                                >< NlaIV
                                >< Eco64I
                                > < RsaI >< BscBI
                                >< Csp6I >< BanI
                                > < AfaI>< AccBII
                                > < NlaIII

```

FIGURE 13.17

31/83

```

TTGGCTCATG TGGTTTATCA TTAGTATTGT ACAAATGGCA CCCGTTTCTG CAATGGTTAG GATGTACATC
  7290      7300      7310      7320      7330      7340      7350

                                TthHB8I ><
                                >< TaqI
                                MnlI ><
                                Ksp632I ><
                                >< FokI
                                >< EarI
                                >< MboII EarI ><
>< FokI
>< Eam1104I>< AluI>< MboII >< NlaIII Eam1104I ><
TTCTTTGCTT CTTTCTACTA CATATGGAAG AGCTATGTTT ATATCATGGA TGGTTGCACC TCTTCGACTT
  7360      7370      7380      7390      7400      7410      7420

                                XhoII ><
                                Sau3AI ><
                                NlaIII ><
                                NdeII ><
                                MflI ><
                                MboI ><
                                >< ThaI
                                >< MvnI
                                >< EarI
>< HinP1I >< MluI
>< Hin6I >< BstUI
>< HhaI >< Bsp50I >< RsaI
>< NlaIII >< CfoI >< AflIII >< Csp6I >< Tru9I BspAI ><
>< BspWI >< BspWI >< AccII >< AfaI >< MseI BglII ><
GCATGATGTG CTATAAGCGC AATCGTGCCA CACGCGTTGA GTGTACAAC ATTGTTAATG GCATGAAGAG
  7430      7440      7450      7460      7470      7480      7490

                                >< Pali
                                >< HaeIII
                                >< DsaI
                                >< MunI
>< MboII >< BsuRI
>< DpnI >< BshI
>< Bsp143I >< MnlI >< BsaJI >< PleI>< HinfI
ATCTTTCTAT GTCTATGCAA ATGGAGGCCG TGGCTTCTGC AAGACTCACA ATTGGAATTG TCTCAATTGT
  7500      7510      7520      7530      7540      7550      7560

                                >< RsaI
                                >< Csp6I
                                >< BsrI
                                >< AfaI
                                >< GsuI
                                >< BpmI
                                >< MaeIIIDraI ><
                                >< BsrI
GACACATTTT GCACTGGTAG TACATTCATT AGTGATGAAG TTGCTCGTGA TTTGTCACCTC CAGTTTAAAA
  7570      7580      7590      7600      7610      7620      7630

                                >< ThaI
                                >< MvnI
                                >< HphI
                                >< HinP1I ><
                                >< HinP1I
                                >< Hin6I
                                >< Hin6I
                                >< HhaI ><
                                >< HhaI
                                >< CfoI ><
                                >< CfoI
                                >< BstUI
                                >< BssHII
                                >< Bsp50I ><
                                >< BsrI
                                >< AccII
GACCAATCAA CCCTACTGAC CAGTCATCGT ATATTGTTGA TAGTGTTGCT GTGAAAAATG GCGCGCTTCA
  7640      7650      7660      7670      7680      7690      7700

```

FIGURE 13. 18

32/83

```

                                >< FokI
                                >< BsmAI
                                >< Alw26I   >< AciI
                                >< MnlI
CCTCTACTTT GACAAGGCTG GTCAAAAGAC CTATGAGAGA CATCCGCTCT CCCATTTTGT CAATTTAGAC
  7710      7720      7730      7740      7750      7760      7770

                                >< VspI
                                >< Tru9I
                                >< MseI
                                >< AsnI
                                >< AseI
                                >< BcgI/a
> < AluI
AATTTGAGAG CTAACAACAC TAAAGGTTCA CTGCCTATTA ATGTCATAGT TTTTGATGGC AAGTCCAAAT
  7780      7790      7800      7810      7820      7830      7840

                                >< SfcI   >< PvuII
                                >< RsaI   >< Psp5I
                                >< Csp6I   >< NspBII
                                >< AfaI   >< AluI
>< HinfI   >< DdeI   >< BcgI   >< AfaI   >< AluI
GCGACGAGTC TGCTTCTAAG TCTGCTTCTG TGTACTACAG TCAGCTGATG TGCCAACCTA TTCTGTTGCT
  7850      7860      7870      7880      7890      7900      7910

                                TthHB8I ><
                                TagI ><
                                SalI ><
                                RtrI ><
                                HindII ><
                                HincII ><
                                >< ScaI
                                >< RsaI   >< Tru9I
                                >< Csp6I   >< SfaNI >< Eco57I
                                >< AfaI   >< MseI   AccI ><
>< AluI   >< MaeII
TGACCAAGCT CTTGTATCAG ACGTTGGAGA TAGTACTGAA GTTTCGGTTA AGATGTTTGA TGCTTATGTC
  7920      7930      7940      7950      7960      7970      7980

                                >< Tru9I
                                >< MseI
                                > < Esp4I   >< SfcI
                                > < AflIII   >< BspWI >< AluI
GACACCTTTT CAGCAACTTT TAGTGTTTCCT ATGGAAAAAC TTAAGGCACT TGTTGCTACA GCTCACAGCG
  7990      8000      8010      8020      8030      8040      8050

                                >< PvuII
                                >< Psp5I
                                >< NspBII
                                >< Fnu4HI
                                >< AluI
                                >< BbvI
>< AluI
AGTTAGCAAA GGGTGTAGCT TTAGATGGTG TCCTTTCTAC ATTCGTGTCA GCTGCCCCGAC AAGGTGTTGT
  8060      8070      8080      8090      8100      8110      8120

                                MaeIII ><
                                >< DdeI
                                >< BsmAI
                                >< BfrI
                                >< FokI >< Alw26I
TGATACCGAT GTTGACACAA AGGATGTTAT TGAATGTCTC AAACCTTCAC ATCACTCTGA CTTAGAAGTG
  8130      8140      8150      8160      8170      8180      8190

                                >< XhoII
                                Sau3AI ><
                                >< NdeII
                                >< MflI
                                >< MboI
                                >< NlaIII >< HgaI
                                >< HinfI >< DpnII
                                DpnI ><

```

FIGURE 13.19

33/83

```

                                Bsp143I ><
                                >< BsaHI >< BstYI
                                >< BbiIII >< BspAI
                                >< AcyI >< BglII
                                >< MaeIII>< HphI
                                >< MaeIII >< HphI >< NlaIII
ACAGGTGACA GTTGTAAACA TTTCATGCTC ACCTATAATA AGGTTGAAAA CATGACGCCC AGAGATCTTG
      8200      8210      8220      8230      8240      8250      8260

                                >< NspI
                                >< NspHI
                                >< NlaIII
>< HinPII
>< Hin6I
  >< HhaI
  >< CfoI
                                >< BspWI >< MaeIII
GCGCATGTAT TGAAGTGAAT GCAAGGCATA TCAATGCCCA AGTAGCAAAA AGTCACAATG TTTCATCAT
      8270      8280      8290      8300      8310      8320      8330

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< Eam1105I
                                >< BbvI
                                >< PvuII
                                >< Psp5I
                                >< NspBII
                                >< Fnu4HI
                                >< AflIII
                                >< AluI >< BbvI >< Fnu4HI
CTGGAATGTA AAAGACTACA TGTCTTTATC TGAACAGCTG CGTAAACAAA TTCGTAGTGC TGCCAAGAAG
      8340      8350      8360      8370      8380      8390      8400

                                >< RmaI
                                >< MaeI >< Eam1105I
                                >< MboII
AACAAACATAC CTTTTAGACT AACTTGTGCT ACAACTAGAC AGGTTGTCAA TGTCATAACT ACTAAAATCT
      8410      8420      8430      8440      8450      8460      8470

                                >< Tru9I
                                >< Pali
                                >< MseI
                                >< HaeIII
                                >< Esp4I
                                >< BsuRI
                                >< BshI
                                >< ScaI
                                >< RsaI >< Tru9I
                                >< Csp6I >< MseI
                                >< AfaI >< DraI >< AflIII >< BbvI
CACTCAAGGG TGGTAAGATT GTTAGTACTT GTTTTAACT TATGCTTAAG GCCACATTAT TGTGCGTTCT
      8480      8490      8500      8510      8520      8530      8540

                                >< RsaI
                                >< Csp6I
                                >< BsrI
                                >< NlaIII
                                >< Fnu4HI
                                >< AfaI
                                >< MaeIII
TGCTGCATTG GTTTGTTATA TCGTTATGCC AGTACATACA TTGTCAATCC ATGATGGTTA CACAAATGAA
      8550      8560      8570      8580      8590      8600      8610

                                >< MaeIII
                                >< MaeIII
                                >< FokI
ATCATTGGTT ACAAAGCCAT TCAGGATGGT GTCACCTCGT ACATCATTTT TACTGATGAT TGTTTTGCAA
      8620      8630      8640      8650      8660      8670      8680

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< HgaI
                                >< BstXI
                                >< BbvI
                                >< AluI
                                SfcI >
                                Fnu4HI ><
                                BbvI ><
ATAAACATGC TGGTTTTGAC GCATGGTTTA GCCAGCGTGG TGGTTCATAC AAAAATGACA AAAGCTGCCC
      8690      8700      8710      8720      8730      8740      8750

```

FIGURE 13. 20

```

                                >< ScrFI
                                >< ScrFI >< RsaI
                                >< MvaI >< MspI
                                >< EcoRII >< HpaII
                                >< Ecl136I>< NciI
                                >< DsaV >< HapII
                                >< BstOI>< DsaV
                                >< BstNI >< Csp6I
                                >< BsiLI >< BcnIDdeI ><
                                >< ApyI >< AfaI
                                >< Fnu4HI
                                >< AluI
TGTAGTAGCT GCTATCATT CAAGAGAGAT TGGTTTCATA GTGCCTGGCT TACCGGGTAC TGTGCTGAGA
8760 8770 8780 8790 8800 8810 8820

                                > < MaeIII >< HphI >< MnlI >< BspWI
GCAATCAATG GTGACTTCTT GCATTTTCTA CCTCGTGTTT TTAGTGCTGT TGGCAACATT TGCTACACAC
8830 8840 8850 8860 8870 8880 8890

                                Tru9I >
                                SfaNI ><
                                >< RsaI
                                MseI >
                                >< BspWI >< Fnu4HI >< Csp6I
                                >< BbvI>< MnlI >< DdeI >< AfaI
CTTCCAAACT CATTGAGTAT AGTGATTTTG CTACCTCTGC TTGCGTTCTT GCTGCTGAGT GTACAATTTT
8900 8910 8920 8930 8940 8950 8960

                                > < RmaI
                                >< MnlI
                                > < MaeI
TAAGGATGCT ATGGGCAAAC CTGTGCCATA TTGTTATGAC ACTAATTGCT TAGAGGGTTC TATTTCTTAT
8970 8980 8990 9000 9010 9020 9030

                                ScrFI >
                                MvaI >
                                MnlI ><
                                EcoRII ><
                                Ecl136I >
                                DsaV ><
                                BstOI >
                                BstNI >
                                >< NlaIV >< FokI >< BsiLI >
                                >< BscBI >< ApyI >
                                >< AluI
AGTGAGCTTC GTCCAGACAC TCGTTATGTG CTTATGGATG GTTCCATCAT ACAGTTTCCT AACACTTACC
9040 9050 9060 9070 9080 9090 9100

                                >< RsaI
                                >< SfcI >< NspI
                                >< ScaI >< NspHI
                                >< RsaI >< NlaIII
                                >< SfaNI >< Csp6I >< NlaIII
                                > < MaeIII >< AfaI >< Csp6I
                                >< GsuI >< AfaI >< Csp6I
                                >< BpmI >< DdeI >< AccI >< AfaI
TGGAGGGTTC TGTTAGAGTA GTAACAACCT TTGATGCTGA GTACTGTAGA CATGGTACAT GCGAAAGGTC
9110 9120 9130 9140 9150 9160 9170

                                >< SstI
                                >< SduI
                                >< SacI
                                NspII ><
                                HgiAI ><
                                Eco24I ><
                                Bsp1286I ><

```

FIGURE 13.21

35/83

```

Ecl1136II ><> BmyI
BanII ><
>< Tru9I Alw21I ><
>< BsrI >< MseI >< AluI
AGAAGTAGGT ATTTGCCTAT CTACCAGTGG TAGATGGGTT CTTAATAATG AGCATTACAG AGCTCTATCA
9180 9190 9200 9210 9220 9230 9240

>< TfiI
>< SfaNI >< HinfI >< AluI >< MnlI
GGAGTTTTCT GTGGTGTGGA TGCGATGAAT CTCATAGCTA ACATCTTTAC TCCTCTTGTG CAACCTGTGG
9250 9260 9270 9280 9290 9300 9310

>< MaeIII
HphI ><
>< Eco57I >< BbvI Fnu4HI ><
GTGCTTTAGA TGTGTCTGCT TCAGTAGTGG CTGGTGGTAT TATTGCCATA TTGGTGACTT GTGCTGCCCTA
9320 9330 9340 9350 9360 9370 9380

>< RsaI
>< Csp6I >< NlaIII
>< MaeII >< BbvI >< Fnu4HI
>< AflIII >< AfaI >< HphI >< BspWI
CTACTTTATG AAATTCAGAC GTGTTTTTGG TGAGTACAAC CATGTTGTTG CTGCTAATGC ACTTTTGTTT
9390 9400 9410 9420 9430 9440 9450

>< RsaI
>< NlaIV
>< KpnI
>< Eco64I >< ScrFI
>< Csp6I >< NciI
>< BscBI >< MspI
>< Asp718 >< HpaII
>< BanI >< AluI >< HinfI
>< AfaI >< HapII >< PleI
>< AccB1I >< BcnI >< DdeI
>< Acc65I >< AluI >< DsaV >< AccI
TTGATGTCTT TCACTATACT CTGTCTGGTA CCAGCTTACA GCTTTCTGCC GGGAGTCTAC TCAGTCTTTT
9460 9470 9480 9490 9500 9510 9520

>< RsaI
>< Csp6I
>< AfaI >< HphI >< HphI NlaIII ><
ACTTGTAATT GACATTCTAT TTCACCAATG ATGTTTCATT CTTGGCTCAC CTTCAATGGT TTGCCATGTT
9530 9540 9550 9560 9570 9580 9590

TTCTCCTATT GTGCCTTTTT GGATAACAGC AATCTATGTA TTCTGTATTT CTCTGAAGCA CTGCCATTGG
9600 9610 9620 9630 9640 9650 9660

>< TthHB8I
>< RsaI
>< MnlI
>< MnlI
>< Csp6I
>< Tru9I >< PleI >< BcgI/a >< TaqI
>< MseI >< DdeI >< NlaIII >< BbvI
>< Eco57I >< BfrI >< HinfI >< MseI >< MaeIII >< AfaI Fnu4HI ><
TTCTTTAACA ACTATCTTAG GAAAAGAGTC ATGTTTAATG GAGTTACATT TAGTACCTTC GAGGAGGCTG
9670 9680 9690 9700 9710 9720 9730

>< RsaI
>< Csp6I
>< BcgI >< RsaI
>< Csp6I >< BsmAI

```

FIGURE 13.22

36/83

```

    >< AfaI          >< AfaI          >< Alw26I
CTTTGTGTAC CTTTTTGCTC AACAAAGGAAA TGTACCTAAA ATTGCGTAGC GAGACACTGT TGCCACTTAC
  9740      9750      9760      9770      9780      9790      9800

                                >< NlaIV
                                >< DdeI
                                >< BscBI
                                >< BfrI   AluI ><
    ACAGTATAAC AGGTATCTTG CTCTATATAA CAAGTACAAG TATTTTCAGTG GAGCCTTAGA TACTACCAGC
      9810      9820      9830      9840      9850      9860      9870

    >< Fnu4HI
    >< DdeI
    >< Fnu4HI    >< BfrI
    >< BbvI    >< AluI    >< BbvI
TATCGTGAAG CAGCTTGCTG CCACTTAGCA AAGGCTCTAA ATGACTTTAG CAACTCAGGT GCTGATGTTC
  9880      9890      9900      9910      9920      9930      9940

                                >< SfcI
                                >< PstI
                                >< BsmI
TCTACCAACC ACCACAGACA TCAATCACTT CTGCTGTTCT GCAGAGTGGT TTTAGGAAAA TGGCATTCCC
  9950      9960      9970      9980      9990      10000      10010

    >< RsaI
    >< NlaIII
    >< MaeIII
    >< Csp6I
    >< AfaI
GTCAGGCAAA GTTGAAGGGT GCATGGTACA AGTAACCTGT GGAAC TACAA CTCTTAATGG ATTGTGGTTG
  10020      10030      10040      10050      10060      10070      10080

                                XhoII ><
                                Sau3AI ><
                                >< Tru9I   NdeII ><
                                >< NspI     MflI ><
                                >< NspHI    MboI ><
                                >< NlaIII   DpnII ><
                                >< MseI     BstYI ><
                                >< MboII   BspAI ><
                                >< BbsI     BglII ><
GATGACACAG TATACTGTCC AAGACATGTC ATTTGCACAG CAGAAGACAT GCTTAATCCT AACTATGAAG
  10090      10100      10110      10120      10130      10140      10150

                                Pali >
                                MscI >
                                HaeIII >
                                EaeI ><
                                BsuRI >
                                BshI >
                                Bali >
    >< DpnI >< MboII
    >< Bsp143I
    >< AluI
ATCTGCTCAT TCGCAAATCC AACCATAGCT TTCTTGTTCA GGCTGGCAAT GTTCAACTTC GTGTTATTGG
  10160      10170      10180      10190      10200      10210      10220

    >< DdeI> < Tru9I
    >< BfrI> < MseI
    >< DdeI
CCATTCTATG CAAAATTGTC TGCTTAGGCT TAAAGTTGAT ACTTCTAACC CTAAGACACC CAAGTATAAA
  10230      10240      10250      10260      10270      10280      10290

    >< ScrFI
    >< MvaI
    >< EcoRII
    >< Ecl136I
                                >< SphI

```

FIGURE 13.23



37/83

```

>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< ApyI
TTTGTCCGTA TCCAACCTGG TCAAACATTT TCAGTTCTAG CATGCTACAA TGGTTCACCA TCTGGTGT TT
10300      10310      10320      10330      10340      10350      10360

>< PaeI
>< NspI
>< NspHI
>< RmaI >< NlaIII
>< MaeI >< HphI

>< Sau3AI
>< NdeII
>< MboI>< NlaIII
>< DpnII
>< Eco31I
>< BsmAI
>< BsaI>< NlaIII
>< Alw26I
ATCAGTGTGC CATGAGACCT AATCATACCA TTAAAGGTTT TTTCCCTTAAT GGATCATGTG GTAGTGTGTTG
10370      10380      10390      10400      10410      10420      10430

>< Zsp2I
>< Ppu10I
>< NsiI>< SfaNI
>< NdeI
>< Mph1103I
>< EcoT22I
>< Tru9I
>< MseI
>< AvaIII
>< AluI
>< RsaI ><
>< Csp6I ><
>< AfaI ><
TTTTAACATT GATTATGATT GCGTGTCTTT CTGCTATATG CATCATATGG AGCTTCCAAC AGGAGTACAC
10440      10450      10460      10470      10480      10490      10500

>< SinI
>< Sau96I
>< NspIV
>< NspHII
>< Eco47I
>< Cfr13I
>< BsiZI
>< Bme18I >< HindII
>< AvaII >< HincII
>< AsuI>< BsgI >< BbvI >< BspMI
>< AfaI ><
GCTGGTACTG ACTTAGAAGG TAAATTCTAT GGTCCATTTG TTGACAGACA AACTGCACAG GCTGCAGGTA
10510      10520      10530      10540      10550      10560      10570

>< Tru9I
>< MseI
>< BbvI
>< Fnu4HI
>< HphI ><
CAGACACAAC CATAACATTA AATGTTTTGG CATGGCTGTA TGCTGCTGTT ATCAATGGTG ATAGGTGGTT
10580      10590      10600      10610      10620      10630      10640

>< Tru9I
>< TfiI
>< MseI
>< HphI
>< HinfI
>< Tru9I
>< MseI
>< RsaI
>< Csp6I
>< AfaI
TCTTAATAGA TTCACCACTA CTTTGAATGA CTTTAACCTT GTGGCAATGA AGTACAACCTA TGAACCTTTG
10650      10660      10670      10680      10690      10700      10710

>< SinI
>< Sau96I
>< PssI
>< Psp5II
>< PpuMI
>< NspIV
>< NspHII
>< NlaIV

```

FIGURE 13. 24

```

                                >< EcoO109I
                                >< Eco47I
                                >< DraII
    >< Sau3AI                    >< Cfr13I
    >< NdeII                     >< BsiZI
    >< MboI                      >< BscBI
    >< DpnII>< NlaIII            >< Bme18I
    >< DpnI >< HindII            >< DdeI
    >< BspAI >< HincII          >< BfrI
    >< Bsp143I                  >< AsuI      >< MnlI      >< BbvI
ACACAAGATC ATGTTGACAT ATTGGGACCT CTTTCTGCTC AAACAGGAAT TGCCGTCTTA GATATGTGTG
10720      10730      10740      10750      10760      10770      10780

                                >< StyI
                                >< RsaI
                                >< EcoT14I
                                >< Eco130I
                                >< SfcI
                                >< Csp6I
    >< Fnu4HI                    >< Fnu4HI      >< BssT1I
    >< BbvI                      >< Fnu4HI      >< BsaJI
    >< BbvI                      >< AluI       >< PstI       >< AfaI
CTGCTTTGAA AGAGCTGCTG CAGAATGGTA TGAATGGTCG TACTATCCTT GGTAGCACTA TTTTAGAAGA
10790      10800      10810      10820      10830      10840      10850

                                >< StyI
                                >< EcoT14I
                                >< Eco130I
                                >< BssT1I
                                >< MaeIII>< BsaJI
    >< MboII
TGAGTTTACA CCATTTGATG TTGTTAGACA ATGCTCTGGT GTTACCTTCC AAGGTAAGTT CAAGAAAATT
10860      10870      10880      10890      10900      10910      10920

                                >< SfaNI
                                >< SduI
                                >< NspII
    >< Tru9I> < Bsp1286I        >< Tru9I
    >< MseI > < BmyI           >< MseI
    GTTAAGGGCA CTCATCATTG GATGCTTTTA ACTTTCTTGA CATCACTATT GATTCTTGTT CAAAGTACAC
10930      10940      10950      10960      10970      10980      10990

                                >< XmnI
                                >< BsmI
                                >< BscCI
    >< MaeIII
AGTGGTCACT GTTTTCTTT GTTTACGAGA ATGCTTTCTT GCCATTTACT CTTGGTATTA TGGCAATTGC
11000      11010      11020      11030      11040      11050      11060

                                >< NspI
                                >< NspHI
                                >< NlaIII
    >< BspWI >< Fnu4HI>< BspWI >< BscCI
TGATGTGCT ATGCTGCTTG TTAAGCATAA GCACGCATTC TTGTGCTTGT TTCTGTTACC TTCTCTTGCA
11070      11080      11090      11100      11110      11120      11130

                                >< SfaNI
                                >< RmaI
                                >< NspI
                                >< NlaIII
                                >< NheI
                                >< MaeI
                                >< BsiBI
                                >< NlaIII
    >< BspWI >< MseI >< AccI> < NspHI>< AluI >< BsaBI >< NlaIII
ACAGTTGCTT ACTTTAATAT GGTCTACATG CCTGCTAGCT GGGTGATGCG TATCATGACA TGGCTTGAAT
11140      11150      11160      11170      11180      11190      11200

```

FIGURE 13.25

39/83

```

                                >< Tru9I
                                >< MseI
    > < RmaI                    > < Esp4I
    > < MaeI                    >< Eco57I
                                >< AluI
                                > < AflIII
                                >< AluI
TGGCTGACAC TAGCTTGTCT GGTATAGGC TTAAGGATTG TGTTATGTAT GCTTCAGCTT TAGTTTTGCT
    11210      11220      11230      11240      11250      11260      11270

                                >< RmaI
                                >< MaeII
                                >< MaeI
                                >< Fnu4HI
    > < NlaIII    >< SfaNI    >< BspHI >< AluI    >< BbvI    >< AflIII
TATTCTCATG ACAGCTCGCA CTGTTTATGA TGATGCTGCT AGACGTGTTT GGACACTGAT GAATGTCATT
    11280      11290      11300      11310      11320      11330      11340

                                >< Sau96I
                                >< Pali
                                >< NspIV
                                >< NlaIII
                                >< HaeIII
                                > < DdeI
                                >< Sau3AI
                                >< NdeII
                                >< MboI
                                >< DpnII
                                >< DpnI
                                >< Bsp143I
                                >< BspAI>< AluI
                                >< BfrI
                                >< AsuI
ACACTTGTTT ACAAAGTCTA CTATGGTAAT GCTTTAGATC AAGCTATTTC CATGTGGGCC TTAGTTATTT
    11350      11360      11370      11380      11390      11400      11410

                                >< RmaI
                                >< NlaIII
                                >< MaeI>< SfcI
                                >< AluI>< AluI
    >< MaeIII    >< MnlI    >< MaeIII
CTGTAACTC TAACTATTCT GGTGTCGTTA CGACTATCAT GTTTTAGCT AGAGCTATAG TGTTTGTGTG
    11420      11430      11440      11450      11460      11470      11480

                                DdeI >
                                >< BsrI
                                >< NlaIII BfrI >
TGTTGAGTAT TACCCATTGT TATTTATTAC TGGCAACACC TTACAGTGTA TCATGCTTGT TTATTGTTTC
    11490      11500      11510      11520      11530      11540      11550

                                >< Pali
                                >< HaeIII
                                >< Fnu4HI
                                >< BsuRI
    >< BbvI    >< Fnu4HI    >< BspWI
    >< BbvI    >< BspWI    >< BshI    >< Eco57I >< MaeIII
TTAGGCTATT GTTGCTGCTG CTACTTTGGC CTTTTCTGTT TACTCAACCG TTACTTCAGG CTTACTCTTG
    11560      11570      11580      11590      11600      11610      11620

                                >< ScrFI
                                >< MvaI
                                >< EcoRII
                                >< Ecl136I
                                >< DsaV
                                >< BstOI
                                >< BstNI
                                >< BsiLI
                                > < BsaJI
                                >< BsaJI
                                >< Eco31I
                                >< BsmAI
                                >< BsaI

```

FIGURE 13.26

```

                >< DrdI >< Alw26I                                >< ApyI      DdeI ><
GTGTTTATGA CTACTTGGTC TCTACACAAG AATTTAGGTA TATGAACTCC CAGGGGCTTT TGCCTCCTAA
  11630      11640      11650      11660      11670      11680      11690

                >< Tru9I
                >< MseI
>< SfaNI          > < HindIII> < Tru9I
  >< MnlI          >< AluI > < MseI > < MnlI          > < NlaIII
GAGTAGTATT GATGCTTTCA AGCTTAACAT TAAGTTGTTG GGTATTGGAG GTAAACCATG TATCAAGGTT
  11700      11710      11720      11730      11740      11750      11760

                >< VneI
                >< SnoI
                  >< SduI
                  >< NspII
                  >< HgiAI
                  >< Bsp1286I
                  >< BmyI >< RsaI
                >< RsaI          >< ApaLI          >< MboII
                >< Csp6I          >< Alw44I          >< Csp6I          DdeI >
                >< AfaI          >< MaeII          >< Alw21I >< AfaI          BfrI >
GCTACTGTAC AGTCTAAAAT GTCTGACGTA AAGTGCACAT CTGTGGTACT GCTCTCGGTT CTTCAACAAC
  11770      11780      11790      11800      11810      11820      11830

                >< NspII> < RsaI
                  >< DraIII
                  >< SduI>< Csp6I
                  >< Bsp1286I
                >< MboII          >< BmyI > < AfaI          >< MboII
                  >< HinfI >< PleI          >< BspWI >< NlaIII
TTAGAGTAGA GTCATCTTCT AAATTGTGGG CACAATGTGT ACAACTCCAC AATGATATTC TTCTTGCAAA
  11840      11850      11860      11870      11880      11890      11900

                >< TthHB8I
                >< TaqI
                  >< HindIII          >< MboII          >< NlaIII
                  >< AluI          > < Eco57I          >< BspWI AccI ><
AGACACAAC TGAAGCTTTCG AGAAGATGGT TTCTCTTTTG TCTGTTTTGC TATCCATGCA GGGTGCTGTA
  11910      11920      11930      11940      11950      11960      11970

                >< VspI
                >< Tru9I
                >< MseI
                >< AsnI
                >< AseI>< MnlI >< BcgI/a          >< Eco57I          >< Eco57I >< BcgI
GACATTAATA GGTGTGCGA GGAAATGCTC GATAACCGTG CTACTCTTCA GGCTATTGCT TCAGAATTTA
  11980      11990      12000      12010      12020      12030      12040

                >< StuI
                >< ScrFI
                  >< Pali
                  >< MvaI>< HaeIII
                >< EcoRII>< Eco147I
                  >< Ecl136I
                >< DsaV >< BsuRI
                  >< BstOI
                  >< BstNI
                  >< BspWI
                  >< BsiLI
                >< Fnu4HI          >< BsaJI >< BshI          TfiI ><
                >< NdeI          >< BspWI>< MnlI >< BglI          >< SfcI HinfI ><
                >< AciI          >< ApyI>< AatI          > < AluI

```

FIGURE 13. 27

41/83

```

GTTCTTTACC ATCATATGCC GCTTATGCCA CTGCCCAGGA GGCCTATGAG CAGGCTGTAG CTAATGGTGA
12050      12060      12070      12080      12090      12100      12110

    >< XmnI          >< Tru9I          >< SfaNI
    >< HphI          >< MseI          >< DdeI
    >< Asp700I       >< Eco57I       >< BbvI Fnu4HI ><
TTCTGAAGTC GTTCTCAAAA AGTTAAAGAA ATCTTTGAAT GTGGCTAAAT CTGAGTTTGA CCGTGATGCT
12120      12130      12140      12150      12160      12170      12180

                                XhoII ><
                                Sau3AI ><
                                NdeII ><
                                MnlI >
                                >< MnlI
                                >< MflI
                                >< MboI
                                DpnII ><
                                DpnI ><
                                DdeI ><
                                BstYI ><
                                >< BspWI
                                >< RsaIBspAI ><
    >< NlaIII       >< BspAI          >< Csp6IBsp143I ><
                                >< Bsp143I       >< AfaIBglIII ><
GCCATGCAAC GCAAGTTGGA AAAGATGGCA GATCAGGCTA TGACCCAAAT GTACAAACAG GCAAGATCTG
12190      12200      12210      12220      12230      12240      12250

                                >< SpeI
                                >< RmaI
                                >< MaeIII
                                >< MaeI
                                >< MboII
                                >< BspWI
                                >< Eam1104I >< BspWI
                                >< EarI>< BfrI >< AluI
AGGACAAGAG GGCAAAAGTA ACTAGTGCTA TGCAAACAAT GCTCTTCACT ATGCTTAGGA AGCTTGATAA
12260      12270      12280      12290      12300      12310      12320

                                >< ThaI
                                >< MvnI
                                >< HinPII
                                >< Hin6I
                                >< HhaI
                                >< CfoI
                                >< BstUI
                                >< Bsp50I
                                >< AccII
                                SfcI ><
TGATGCACTT AACAAACATTA TCAACAATGC GCGTGATGGT TGTGTTCCAC TCAACATCAT ACCATTGACT
12330      12340      12350      12360      12370      12380      12390

                                >< RsaI
                                >< NlaIV
                                >< Eco64I
                                >< Csp6I
                                >< BslI
                                >< BsiYI>< KpnI
                                >< BscBI
                                >< BanI
                                >< Asp718
                                >< AfaI
                                >< AccB1I
                                >< Acc65I
                                >< MaeIII
                                >< BsgI ><
ACAGCAGCCA AACTCATGGT TGTGTGCCCT GATTATGGTA CCTACAAGAA CACTTGTGAT GGTAACACCT
12400      12410      12420      12430      12440      12450      12460

                                >< Zsp2I
                                >< Ppu10I

```

FIGURE 13.28

42/83

```

    >< NsiI
    >< Mph1103I
    >< NdeI>< EcoT22I
    >< AvaIII >< SfaNI
    >< SfaNI
    >< AciI
    DdeI ><
    BfrI ><
    TTACATATGC ATCTGCACTC TGGGAAATCC AGCAAGTTGT TGATGCGGAT AGCAAGATTG TTCAACTTAG
    12470      12480      12490      12500      12510      12520      12530

    >< Pali
    >< HaeIII
    >< MnlI
    >< DdeIDdeI ><
    >< BsuRI
    >< MaeIII
    >< BspWI
    >< Tru9I>< NlaIII
    >< MseI>< HphI
    >< XcmI>< BshI
    >< AluI
    BspWI ><
    TGAAATTAAC ATGGACAATT CACCAAATTT GGCTTGGCCT CTTATTGTGA CAGCTCTAAG AGCCAACTCA
    12540      12550      12560      12570      12580      12590      12600

    RsaI ><
    NlaIV ><
    KpnI ><
    >< Fnu4HI
    Eco64I ><
    Csp6I ><
    BscBI ><
    Asp718 ><
    AfaI ><
    >< AciI>< BanI
    AccB1I ><
    >< PshAI
    Acc65I ><
    >< Tru9I
    >< PvuII
    >< Psp5I
    >< NspBII
    >< MseI
    >< HinfI >< P1eI
    >< DdeI>< BsrI
    >< PshAI
    Acc65I ><
    >< AluI >< SfcI
    GCTGTAAAC TACAGAATAA TGAAGTGAAGT CCAGTAGCAC TACGACAGAT GTCCTGTGCG GCTGGTACCA
    12610      12620      12630      12640      12650      12660      12670

    >< TthHB8I
    >< TaqI
    >< SfuI
    >< NspV
    >< MnlI
    >< LspI
    >< Csp45I
    >< BstBI
    >< Bsp119I
    >< BsiCI
    >< Bpu14I
    >< AsuII
    >< RsaI
    >< Csp6I
    >< AluI
    >< AfaI
    CACAAACAGC TTGTACTGAT GACAATGCAC TTGCCTACTA TAACAATTCTG AAGGGAGGTA GGTTTGTGCT
    12680      12690      12700      12710      12720      12730      12740

    >< XhoII
    >< Sau3AI
    >< NdeII
    >< MflI
    >< MboI
    >< DpnII
    >< DpnI
    >< BstYI
    >< BspAI
    >< Bsp143I
    >< BglII
    >< TfiI
    >< RmaI
    >< HinfI
    >< MaeI
    >< DdeI
    >< RsaI
    >< Csp6I
    >< Csp6I>< RsaI
    >< AfaI>< AfaI
    GGCATTACTA TCAGACCACC AAGATCTCAA ATGGGCTAGA TTCCCTAAGA GTGATGGTAC AGGTACAATT
    12750      12760      12770      12780      12790      12800      12810

    >< Sau96I
    >< PssI
    >< Pali
    >< NspIV

```

FIGURE 13.29

```

                                >< HaeIII
                                >< EcoO109I
                                >< DraII
                                >< Cfr13I
                                >< BsuRI
                                >< BsiZI
                                >< BshI
                                >< AsuI
                                RsaI >
                                Csp6I ><
                                AfaI >
TACACAGAAC TGGAACCACC TTGTAGGTTT GTTACAGACA CACCAAAAGG GCCTAAAGTG AAATACTTGT
12820      12830      12840      12850      12860      12870      12880

                                >< SfcI
                                > < MboII
                                MaeII ><
                                >< Fnu4HI >< RsaI
                                >< Eco57I >< Csp6I
                                > < BbsI
                                >< Tru9I
                                >< MseI >< MnlI
                                >< BbvI
                                >< AluI
                                >< AfaI
ACTTCATCAA AGGCTTAAAC AACCTAAATA GAGGTATGGT GCTGGGCAGT TTAGCTGCTA CAGTACGTCT
12890      12900      12910      12920      12930      12940      12950

                                >< RsaI
                                >< SfcI >< Csp6I
                                >< BspWI >< AfaI >< BspMI
                                >< AccI ><
TCAGGCTGGA AATGCTACAG AAGTACCTGC CAATTCAACT GTGCTTTCCT TCTGTGCTTT TGCAGTAGAC
12960      12970      12980      12990      13000      13010      13020

                                >< RmaI
                                >< MnlI
                                >< MaeI >< HphI
CCTGCTAAAG CATATAAGGA TTACCTAGCA AGTGGAGGAC AACCAATCAC CAACTGTGTG AAGATGTTGT
13030      13040      13050      13060      13070      13080      13090

                                >< SinI
                                >< Sau96I
                                >< NspIV
                                >< NspHII
                                >< NlaIII
                                >< Eco47I
                                >< Eam1105I
                                >< Cfr13I
                                >< BsiZI
                                >< Bme18I >< XcmI
                                >< AvaII PleI ><
                                >< MaeIII >< AluI >< AsuI >< HinfI
GTACACACAC TGGTACAGGA CAGGCAATTA CTGTAACACC AGAAGCTAAC ATGGACCAAG AGTCCTTTGG
13100      13110      13120      13130      13140      13150      13160

                                >< TfiI
                                >< MaeIII
                                >< SfaNI
                                >< NlaIII >< FokI
                                >< HinfI
TGGTGCTTCA TGTGTCTGT ATTGTAGATG CCACATTGAC CATCCAAATC CTAAAGGATT CTGTGACTTG
13170      13180      13190      13200      13210      13220      13230

                                > < RsaI
                                >< MaeII
                                >< Csp6I
                                > < AfaI
                                >< BsrI
                                >< DdeI
                                >< BfrI
AAAGGTAAGT ACGTCCAAAT ACCTACCACT TGTGCTAATG ACCCAGTGGG TTTTACACTT AGAAACACAG
13240      13250      13260      13270      13280      13290      13300

                                >< ThaI

```

FIGURE 13.30

```

>< RsaI
>< Csp6I
>< AfaI >< AciI
TCTGTACCGT CTGCGGAATG TGGAAAGGTT ATGGCTGTAG TTGTGACCAA CTCCGCGAAC CCTTGATGCA
13310 13320 13330 13340 13350 13360 13370

>< SfaNI
>< MvnI
>< BstUI
>< Bsp50I
>< AciI
>< SfcI >< MaeIII >< AccIISfaNI ><
TCTGTACCGT CTGCGGAATG TGGAAAGGTT ATGGCTGTAG TTGTGACCAA CTCCGCGAAC CCTTGATGCA
13310 13320 13330 13340 13350 13360 13370

>< Zsp2I
> < SfaNI
>< Mph1103I>< Tru9I
>< Ppu10I>< MaeII Fnu4HI ><
>< NsiI> < FokI BsgI ><
>< EcoT22I >< MseI >< BbvI
>< AciI>< AvaIII >< DraI >< AciI >< Fnu4HI AciI ><
GTCTGCGGAT GCATCAACGT TTTTAAACGG GTTTGCGGTG TAAGTGCAGC CCGTCTTACA CCGTGC GGCA
13380 13390 13400 13410 13420 13430 13440

>< SpeI
>< ScaI
>< RsaI
>< RmaI
>< MaeI
> < Csp6I >< SfcI >< BspWI
>< BspWI >< AfaI >< AccI >< BcgI/a BcgI '
CAGGCACTAG TACTGATGTC GTCTACAGGG CTTTGTATAT TTACAACGAA AAAGTTGCTG GTTTTGCAAA
13450 13460 13470 13480 13490 13500 13510

>< ScrFI
>< MvaI
>< MnlI
>< EcoRII
>< Ecl136I
>< BstOI
>< BstNI
>< BslI
>< DsaV >< BsiYI
>< BsiLI >< PleI
>< ApyI > < FokI >< HinfI
GTTCTTAAAA ACTAATTGCT GTCGCTTCCA GGAGAAGGAT GAGGAAGGCA ATTTATTAGA CTCTTACTTT
13520 13530 13540 13550 13560 13570 13580

>< NlaIII
>< Ksp632I
>< EarI
>< Eam1104I
>< BsmAI >< Tru9I
>< MnlI >< Alw26I >< MboII >< MseI
GTAGTTAAGA GGCATACTAT GTCTAACTAC CAACATGAAG AGACTATTTA TAACTTGGTT AAAGATTGTC
13590 13600 13610 13620 13630 13640 13650

>< RsaI
>< NlaIV
> < NlaIII
>< KpnI
>< HphI
> < Eco64I
>< Csp6I
>< BscBI
> < BanI
> < Asp718

```

FIGURE 13.31



```

                                >< MaeIII >< AfaI
                                > < AccB1I MaeII ><
>< NspBII                                >< Acc65I > < HgaI
>< AciI                                >< NlaIII
CAGCGGTTGC TGTCCATGAC TTTTCAAGT TTAGAGTAGA TGGTGACATG GTACCACATA TATCACGTCA
13660      13670      13680      13690      13700      13710      13720

                                >< MnlI
                                >< MaeII
GCGTCTAACT AAATACACAA TGGCTGATTT AGTCTATGCT CTACGTCATT TTGATGAGGG TAATTGTGAT
13730      13740      13750      13760      13770      13780      13790

>< Tru9I
>< MseI                                >< MaeIII >< MunI
ACATTAAGAG AAATACTCGT CACATACAAT TGCTGTGATG ATGATTATTT CAATAAGAAG GATTGGTATG
13800      13810      13820      13830      13840      13850      13860

                                >< ThaI
                                >< MvnI
                                >< MluI
                                >< BstUI
                                >< Bsp50I
                                >< RsaI
                                >< HphI
>< TfiI                                >< AflIII >< DdeI >< Csp6I Tru9I ><
>< HinfI                                >< AccII >< BfrI >< AfaI MseI ><
ACTTCGTAGA GAATCCTGAC ATCTTACGCG TATATGCTAA CTTAGGTGAG CGTGTACGCC AATCATTATT
13870      13880      13890      13900      13910      13920      13930

                                >< XhoII >
                                >< Sau3AI >
                                >< NdeII >
                                >< MflI >
                                >< MboI >
> < SfaNI                                >< RsaI
>< RsaI                                > < Csp6I
>< Csp6I                                >< BspWI
>< AfaI                                >< AfaI
AAAGACTGTA CAATTCTGCG ATGCTATGCG TGATGCAGGC ATTGTAGGCG TACTGACATT AGATAATCAG
13940      13950      13960      13970      13980      13990      14000

                                > < ScrFI
                                > < MvaI
                                >< Fnu4HI
                                >< EcoRII
                                > < Ecl136I
                                > < BstOI
                                > < BstNI
>< Tru9I                                >< RsaI
>< MseI                                >< HphI
>< DpnI                                >< Csp6I
>< Bsp143I >< BsrI >< BbvI >< ApyI
>< AlwI >< AfaI >< AfaI >< DsaV >< AciI
GATCTTAATG GGAAGTGGTA CGATTTCCGT GATTTTCGTAC AAGTAGCACC AGGCTGCGGA GTTCCTATTG
14010      14020      14030      14040      14050      14060      14070

                                >< SfaNI
                                >< RmaI
                                > < HinfI
>< TfiI >< SfaNI >< MamI >< MnlI >< Fnu4HIPleI ><
>< HinfI >< FokI >< BsiBI >< MaeI >< DdeI
TGGATTCATA TTAATCATTG CTGATGCCCA TCCTCACTTT GACTAGGGCA TTGGCTGCTG AGTCCCATAT
14080      14090      14100      14110      14120      14130      14140

>< Sau3AI
>< NdeII

```

FIGURE 13.32

```

>< MboI
>< MamI
>< DpnII
>< DpnI
>< BspWI
>< BspAI
>< Bsp143I
>< BsiBI
>< BsaBI >< FokI
GGATGCTGAT CTCGCAAAAC CACTTATTAA GTGGGATTTG CTGAAATATG ATTTTACGGA AGAGAGACTT
14150      14160      14170      14180      14190      14200      14210

>< XcmI
>< Tru9I
>< MseI
>< SinI
>< Sau96I
>< NspIV
>< NspHII
>< NlaIV
>< FokI
>< Eco47I
>< Cfr13I
>< BsiZI
>< SspI>< BscBI
>< BsmAI >< Tru9I >< Bme18I
>< MboII >< BsiEI> < MseI >< AvaII
>< Alw26I >< DraI >< AsuI
TGTCTCTTCG ACCGTTATTT TAAATATTGG GACCAGACAT ACCATCCCAA TTGTATTAAC TGTTTGGATG
14220      14230      14240      14250      14260      14270      14280

>< Tru9I
>< MseI
>< FokI
>< MseI
ATAGGTGTAT CCTTCATTGT GCAAACTTTA ATGTGTTATT TTCTACTGTG TTTCCACCTA CAAGTTTGGG
14290      14300      14310      14320      14330      14340      14350

>< SpeI
>< RmaI
>< MaeI >< SspI
ACCACTAGTA AGAAAAATAT TTGTAGATGG TGTTCCTTTT GTTGTTTCAA CTGGATACCA TTTTCGTGAG
14360      14370      14380      14390      14400      14410      14420

>< BsrI
>< ThaI>< Esp3I
>< DdeI
>< BstUI
>< Bsp50I >< BsmBI
>< MvnI>< BsmAI
>< HgaI>< AluI >< Alw26I
>< FokI >< AccII >< BbvI
TTAGGAGTCG TACATAATCA GGATGTAAAC TTACATAGCT CGCGTCTCAG TTTCAAGGAA CTTTTAGTGT
14430      14440      14450      14460      14470      14480      14490

>< Zsp2I
>< SphI
>< Ppu10I
>< PaeI
>< NspI

```

FIGURE 13.33

47/83

```

>< Sau3AI          >< NspHI
>< NdeII           >< NsiI
>< MboI            >< NlaIII
>< DpnII           >< Mph1103I
> < DpnI          >< Fnu4HI
>< Fnu4HI>< BspWI  >< EcoT22I
>< BspAI           >< BspWI
> < Bsp143I> < AvaIII > < AlwNI
>< AlwI            >< AluI   >< AluI   >< BbvI   >< MaeI
ATGCTGCTGA TCCAGCTATG CATGCAGCTT CTGGCAATTT ATTGCTAGAT AAACGCACTA CATGCTTTTC
14500      14510      14520      14530      14540      14550      14560

>< ScrFI
>< NciI
>< MspI
>< HpaII
>< HapII
>< Fnu4HI
>< AlwNI
>< AluI
AGTAGCTGCA CTAACAAACA ATGTTGCTTT TCAAAGTCTC AAACCCGGTA ATTTTAATAA AGACTTTTAT
14570      14580      14590      14600      14610      14620      14630

>< Tru9I
>< MseI
GACTTTGCTG TGTCTAAAGG TTTCTTTAAG GAAGGAAGTT CTGTTGAAGT AAAACACTTC TTCTTTGCTC
14640      14650      14660      14670      14680      14690      14700

>< FokI
>< Fnu4HI
AGGATGGCAA CGCTGCTATC AGTGATTATG ACTATTATCG TTATAATCTG CCAACAATGT GTGATATCAG
14710      14720      14730      14740      14750      14760      14770

>< VspI
>< Tru9I
>< MseI
>< AsnI
>< AseI
ACAACTCCTA TTCGTAGTTG AAGTTGTTGA TAAATACTTT GATTGTTACG ATGGTGGCTG TATTAATGCC
14780      14790      14800      14810      14820      14830      14840

>< Tru9I
>< MseI
>< HpaI
>< HindII
>< HincII
AACCAAGTAA TCGTTAACAA TCTGGATAAA TCAGCTGGTT TCCCATTAA TAAATGGGGT AAGGCTAGAC
14850      14860      14870      14880      14890      14900      14910

>< SfaNI
>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI
>< PleI
>< HinfI>< MnlI
>< BspAI
>< AlwI
>< SfaNI
>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI
>< Bsp143I
>< BspAI
>< AlwI
>< ThaI
>< MvnI
>< BstUI
>< Bst1107I
>< BspWI
>< FokI
>< Bsp50I
>< AccII>< DdeI
>< AccI
TTTATTATGA CTCAATGAGT TATGAGGATC AAGATGCACT TTTCGCGTAT ACTAAGCGTA ATGTCATCCC
14920      14930      14940      14950      14960      14970      14980

>< SstI
>< SduI
>< SacI

```

FIGURE 13.34

```

                                >< NspII
                                >< HgiAI
                                >< Eco24I
                                > < Ecl136II
                                >< Bsp1286I
                                >< BmyI
                                >< BanII
                                >< Alw21I
                                >< AluI
TACTATAACT CAAATGAATC TTAAGTATGC CATTAGTGCA AAGAATAGAG CTCGCACCGT AGCTGGTGTC
14990      15000      15010      15020      15030      15040      15050

                                >< Tru9I
                                >< TfiI
                                >< MseI
                                >< HinfI
                                > < Esp4I
                                > < AflII
                                >< BspWI
                                > < AluI
TCTATCTGTA G TACTATGAC AAATAGACAG TTTCATCAGA AATTATTGAA GTCAATAGCC GCCACTAGAG
15060      15070      15080      15090      15100      15110      15120

                                >< Tru9I
                                >< MseI
GAGCTACTGT GGTAATTGGA ACAAGCAAGT TTTACGGTGG CTGGCATAAT ATGTTAAAAA CTGTTTACAG
15130      15140      15150      15160      15170      15180      15190

                                >< NspI ><
                                >< NspHI ><
                                >< NlaIII ><
                                >< NlaIII
                                >< DdeI ><
                                >< BspWI ><
                                >< MaeIII
                                >< BfrI ><
TGATGTAGAA ACTCCACACC TTATGGGTTG GGATTATCCA AAATGTGACA GAGCCATGCC TAACATGCTT
15200      15210      15220      15230      15240      15250      15260

                                > < Pali
                                > < HaeIII
                                > < BsuRI
                                > < BshI
                                >< MnlI
                                >< MaeIII
                                >< SfcI ><
AGGATAATGG CCTCTCTTGT TCTTGCTCGC AAACATAACA CTTGCTGTAA CTTATCACAC CGTTTCTACA
15270      15280      15290      15300      15310      15320      15330

                                Tru9I ><
                                ScrFI >
                                MvaI >
                                >< MseI
                                FokI ><
                                EcoRII ><
                                Ecl136I >
                                DsaV ><
                                BstOI >
                                BstNI >
                                >< NlaIII
                                > < Fnu4HI
                                >< AciI
                                >< ApyI >
GGTTAGCTAA CGAGTGTGCG CAAGTATTAA GTGAGATGGT CATGTGTGGC GGCTCACTAT ATGTTAAACC
15340      15350      15360      15370      15380      15390      15400

                                > < SfaNI
                                >< MspI
                                >< HpaII
                                >< HapII
                                >< HphI
                                >< BspWI
                                >< Tru9I
                                >< MaeIII ><
                                >< MseI
                                >< AluI ><

```

FIGURE 13.35

```

AGGTGGAACA TCATCCGGTG ATGCTACAAC TGCTTATGCT AATAGTGTCT TTAACATTTG TCAAGCTGTT
15410      15420      15430      15440      15450      15460      15470

>< BspWI                                     >< AluI       > < AciI
ACAGCCAATG TAAATGCACT TCTTTCAACT GATGGTAATA AGATAGCTGA CAAGTATGTC CGCAATCTAC
15480      15490      15500      15510      15520      15530      15540

                                     >< DrdI
                                     >< Sau3AI
                                     >< NdeII
                                     >< MboI
                                     > < MamI
                                     >< FbaI
                                     >< DpnII
                                     >< DpnI
                                     >< BspHI
                                     >< BspAI
                                     >< Bsp143I
                                     >< BsiQI
                                     >< SfcI
                                     > < BsiBI>< NlaIII
                                     >< BsmAI
                                     > < BsaBI>< FokI
                                     >< Alw26I
                                     >< BclI>< EcoRI
AACACAGGCT CTATGAGTGT CTCTATAGAA ATAGGGATGT TGATCATGAA TTCGTGGATG AGTTTTACGC
15550      15560      15570      15580      15590      15600      15610

                                     >< TfiI
                                     >< SfaNI
                                     >< NlaIII
                                     >< HinfI
                                     >< BspMI
                                     >< MaeIII
TTACCTGCGT AAACATTTCT CCATGATGAT TCTTTCTGAT GATGCCGTTG TGTGCTATAA CAGTAACTAT
15620      15630      15640      15650      15660      15670      15680

                                     > < RmaI
                                     >< NheI >< Tru9I
>< Fnu4HI                                     > < MaeI       >< Tru9I
>< AciI                                     >< AluI >< MseI   >< MseI
GCGGCTCAAG GTTTAGTAGC TAGCATTAAG AACTTTAAGG CAGTTCTTTA TTATCAAAAT AATGTGTTCA
15690      15700      15710      15720      15730      15740      15750

                                     >< SinI
                                     >< Sau96I
                                     >< PssI
                                     >< Psp5II
                                     >< PpuMI
                                     >< NspIV
                                     >< NspHII
                                     >< EcoO109I
                                     >< Eco47I
                                     >< DraII
                                     >< Cfr13I
                                     >< BsiZI
                                     >< Bme18I
                                     >< AvaII
                                     >< AsuI       >< MnlI
>< NlaIII                                     >< BsmAI
>< DdeI                                     >< Alw26I
TGTCTGAGGC AAAATGTTGG ACTGAGACTG ACCTTACTAA AGGACCTCAC GAATTTTGCT CACAGCATAC
15760      15770      15780      15790      15800      15810      15820

                                     >< XhoII
                                     >< Sau3AI
                                     >< NdeII
                                     >< MflI
                                     >< MboI

```

FIGURE 13. 36

50/83

```

                >< RsaI          >< DpnII
            >< MaeII          >< DpnI      > < SspI
                >< Csp6I          >< BstYI      HinPII ><
            >< RmaI          >< BsaAI          >< BspMI      Hin6I ><
            >< MaeI          >< AflIII          >< BspAI      HhaI ><
>< BspWI>< MseI          >< AfaI      >< AlwI>< Bsp143I      CfoI ><
AATGCTAGTT AAACAAGGAG ATGATTACGT GTACCTGCCT TACCCAGATC CATCAAGAAT ATTAGGCGCA
    15830      15840      15850      15860      15870      15880      15890

                >< RsaI          >< SfaNI
            >< TthHB8I          >< Csp6I          >< MaeIII
            >< TaqI          >< AfaI          BsrI ><
GGCTGTTTTG TCGATGATAT TGTCAAAACA GATGGTACAC TTATGATTGA AAGGTTCTGT TCACTGGCTA
    15900      15910      15920      15930      15940      15950      15960

> < FokI
>< BspWI
TTGATGCTTA CCCACTTACA AAACATCCTA ATCAGGAGTA TGCTGATGTC TTCACTTGT ATTTACAATA
    15970      15980      15990      16000      16010      16020      16030

                >< Van91I
                >< PflMI
                >< NspI
            > < Pali>< NspHI
            > < MscI>< NlaIII
            > < HaeIII
            > < BsuRI
            >< BsrI
            >< EaeI      >< BslI      >< NspI
            > < BshI>< BsiYI      >< NspHI
                >< NlaIII          >< AflIII >< AflIII
            >< MaeIII          >< AluI > < BalI>< AccB7I >< NlaIII
CATTAGAAAG TTACATGATG AGCTTACTGG CCACATGTTG GACATGTATT CCGTAATGCT AACTAATGAT
    16040      16050      16060      16070      16080      16090      16100

            >< RsaI> < NlaIV
            >< MnlI
            >< Csp6I      >< DdeI          >< RsaI
                >< BsrI >< MnlI          >< Csp6I
            >< AfaI> < BscBI          >< AfaI          SfcI ><
AACACCTCAC GGTACTGGGA ACCTGAGTTT TATGAGGCTA TGTACACACC ACATACAGTC TTGCAGGCTG
    16110      16120      16130      16140      16150      16160      16170

                >< NlaIV
                >< EcoNI
                >< Eco31I
            >< Eco64I>< BsmAI
            >< BscBI >< BslI
            >< BanI      >< BsiYI
            >< AciI      >< BsaI
            >< BspWI          >< AccB1I>< Alw26I      BbvI ><
TAGGTGCTTG TGTATTGTGC AATTCACAGA CTTCACTTCG TTGCGGTGCC TGTATTAGGA GACCATTCTT
    16180      16190      16200      16210      16220      16230      16240

                >< Tth111I
            >< Fnu4HI          >< NlaIII          > < Tru9I
            >< BspWI >< AspI          > < MseI
ATGTTGCAAG TGCTGCTATG ACCATGTCAT TTCAACATCA CACAAATTAG TGTTGTCTGT TAATCCCTAT
    16250      16260      16270      16280      16290      16300      16310

            >< ScrFI
            >< MvaI

```

FIGURE 13.37

51/83

```

>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< BsaJI
>< ApyI
GTTTGCAATG CCCCAGGTTG TGATGTCAC TATGTGACAC AACTGTATCT AGGAGGTATG AGCTATTATT
16320 16330 16340 16350 16360 16370 16380

>< MaeIII >< MaeIII
>< MnlI
>< MaeI
>< AluI
GCAAGTCACA TAAGCCTCCC ATTAGTTTTT CATTATGTGC TAATGGTCAG GTTTTTGGTT TATACAAAAA
16390 16400 16410 16420 16430 16440 16450

>< NspI
>< NspHI
>< NlaIII>< MaeIII>< MaeIII
>< AflIII
CACATGTGTA GGCAGTGACA ATGTCAC TTTCAATGCG ATAGCAACAT GTGATTGGAC TAATGCTGGC
16460 16470 16480 16490 16500 16510 16520

>< RsaI
>< PleI
>< DdeI
>< Csp6I
>< BsmAI >< HinfI
>< Alw26I >< HindIII
>< AfaI >< AluI >< Fnu4HI >< BbvI
GATTACATAC TTGCCAACAC TTGTACTGAG AGACTCAAGC TTTTCGCAGC AGAAACGCTC AAAGCCACTG
16530 16540 16550 16560 16570 16580 16590

>< ThaI
>< ScaI
>< RsaI >< RsaI
>< MvnI
>< Csp6I >< Csp6I
>< BstUI
>< Bsp50I
>< Tru9I
>< MseI >< NdeI
>< AfaI >< AfaI
>< AluI >< AccII
AGGAAACATT TAAGCTGTCA TATGGTATTG CCACTGTACG CGAAGTACTC TCTGACAGAG AATTGCATCT
16600 16610 16620 16630 16640 16650 16660

MaeIII ><
>< MaeIII
>< EcoO65I
>< Eco91I
>< BstPI
>< BstEII
>< BsrI
TTCATGGGAG GTTGGAACAC CTAGACCACC ATTGAACAGA AACTATGTCT TTAGTGGTTA CCGTGTAAC T
16670 16680 16690 16700 16710 16720 16730

RsaI ><
>< MnlI
>< HphI
>< RsaI
>< Csp6I
>< AfaI
>< AfaI
>< SfaNI
>< MaeIII
>< HphI AfaI ><
AAAAATAGTA AAGTACAGAT TGGAGAGTAC ACCTTTGAAA AAGGTGACTA TGGTGATGCT GTTGTGTACA
16740 16750 16760 16770 16780 16790 16800

```

FIGURE 13. 38

52/83

```

>< RsaI
>< Csp6I
>< AfaI
GAGGTACTAC GACATACAAG TTGAATGTTG GTGATTACTT TGTGTTGACA TCTCACACTG TAATGCCACT
16810      16820      16830      16840      16850      16860      16870

>< VneI
>< SnoI
>< SduI
>< NspII
>< HgiAI
>< DraIII
>< Bsp1286I
>< BmyI
>< ApaLI
>< Alw44I
>< Alw21I
TAGTGCACCT ACTCTAGTGC CACAAGAGCA CTATGTGAGA ATTACTGGCT TGTACCCAAC ACTCAACATC
16880      16890      16900      16910      16920      16930      16940

>< HphI
>< HindII
>< HincII
DdeI ><
BfrI ><

>< SduI
>< NspII
>< HgiAI
>< BspWI
>< DraIII
>< RsaI
>< Csp6I
>< BsrI
>< AfaI
DdeI >
StyI ><
SinI >
Sau96I >
NspIV >
EcoT14I ><
Eco47I >
Eco130I ><
>< ScaI Cfr13I >
BssT1I ><
>< SphI >< RsaI BsiZIF >
>< PaeI BsaJI ><
>< NlaIII Bme18I >
>< NspI>< Csp6I AvaII >
>< NspHI>< AfaI AsuI >
TCAGATGAGT TTTCTAGCAA TGTTGCAAAT TATCAAAAAGG TCGGCATGCA AAAGTACTCT ACACTCCAAG
16950      16960      16970      16980      16990      17000      17010

>< ScrFI
>< RsaI
>< MvaI
>< EcoRII
>< Ecl136I
>< Csp6I
>< BstOI
>< BstNI
>< XcmI
>< BslI
>< NspHII
>< BsiYI
>< BsiLI
>< ApyI
>< BsrI
>< DsaV>< AfaI >< HinfI>< PleI
GACCACCTGG TACTGGTAAG AGTCATTTTG CCATCGGACT TGCTCTCTAT TACCCATCTG CTCGCATAGT
17020      17030      17040      17050      17060      17070      17080

>< SfaNI
>< SphI
>< PaeI
>< NspI
>< NspHI
>< Bst1107I
>< AccI
GTATACGGCA TGCTCTCATG CAGCTGTTGA TGCCCTATGT GAAAAGGCAT TAAAATATTT GCCCATAGAT
17090      17100      17110      17120      17130      17140      17150

>< PvuII
>< Psp5I
>< NspBII
>< Fnu4HI
>< NlaIII>< BspWI
>< AluI
>< BbvI
>< MseI
>< Tru9I
>< SspI

```

FIGURE 13.39



53/83

```

> < Thai
>< Thai
> < MvnI
>< MvnI >< Thai
> < HinPII
>< HinPII
>< HinPII >< MvnI
> < Hin6I
>< Hin6I
> < HhaI
>< HhaI >< HhaI
> < CfoI
>< CfoI >< CfoI
> < BstUI
>< BstUI >< BstUI
>< BssHII
>< BspMI
> < Bsp50I
>< Bsp50I>< Bsp50I
>< TfiI >< Hin6I> < AccII
>< HinfI >< AccII >< AccII
>< EcoRI
AAATGTAGTA GAATCATACC TGCGCGTGCG CGCGTAGAGT GTTTTGATAA ATTCAAAGTG AATTCAACAC
17160 17170 17180 17190 17200 17210 17220

>< Zsp2I
>< Ppu10I
>< NsiI
>< Mph1103I
>< EcoT22I
>< BsgI > < AvaIII >< DrdI
TAGAACAGTA TGTTTTCTGC ACTGTAAATG CATTGCCAGA AACAAGCTGCT GACATTGTAG TCTTTGATGA
17230 17240 17250 17260 17270 17280 17290

>< RmaI
>< MaeI >< MaeII
AATCTCTATG GCTACTAATT ATGACTTGAG TGTTGTCAAT GCTAGACTTC GTGCAAAACA CTACGTCTAT
17300 17310 17320 17330 17340 17350 17360

>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI
>< BspAI
>< AlwI>< Bsp143I > < AciI >< RmaI
>< MaeI SspI ><
ATTGGCGATC CTGCTCAATT ACCAGCCCCC CGCACATTGC TGAATAAGG CACACTAGAA CCAGAAATATT
17370 17380 17390 17400 17410 17420 17430

>< SniI
>< Sau96I
>< NspIV >< StyI
>< NspHII >< NspI
>< Eco47I >< NspHI
>< Cfr13I >< NlaIII
>< BsiZI >< EcoT14I
>< BsgI >< Eco130I
>< Bme18I >< BssT1I
>< AvaII >< BsaJI
>< Tru9I
>< MseI
>< AsuI> < AflIII
TTAATTTCAGT GTGCAGACTT ATGAAAACAA TAGGTCCAGA CATGTTTCCTT GGAAGTTGTC GCCGTTGTCC
17440 17450 17460 17470 17480 17490 17500

```

FIGURE 13. 40

```

      >< HindII
      >< HincII
TGCTGAAATT GTTGACACTG TGAGTGCTTT AGTTTATGAC AATAAGCTAA AAGCACACAA GGATAAGTCA
17510      17520      17530      17540      17550      17560      17570
      >< AluI
GCTCAATGCT TCAAAATGTT CTACAAAGGT GTTATTACAC ATGATGTTTC ATCTGCAATC AACAGACCTC
17580      17590      17600      17610      17620      17630      17640
      >< NlaIII
      >< MnlI
>< EcoNI
      >< BslI
      >< BsiYI
AAATAGGCGT TGTAAGAGAA TTTCTTACAC GCAATCCTGC TTGGAGAAAA GCTGTTTTTA TCTCACCTTA
17650      17660      17670      17680      17690      17700      17710
      >< HphI
      >< AluI
      >< SfcI
      >< DdeI
      >< TfiI
      >< BfrI
      >< HinfI
TAATTCACAG AACGCTGTAG CTTCAAAAAT CTTAGGATTG CCTACGCAGA CTGTTGATTTC ATCACAGGGT
17720      17730      17740      17750      17760      17770      17780
      >< Tth111I
      >< AspI
      >< HindII
      >< HincII
      >< AciI
TCTGAATATG ACTATGTCAT ATTCACACAA ACTACTGAAA CAGCACACTC TTGTAATGTC AACCGCTTCA
17790      17800      17810      17820      17830      17840      17850
      >< XhoII
      >< Sau3AI
      >< NdeII
      >< MflI
      >< MboI
      >< MamI
      >< DpnII
      >< DpnI
      >< BstYI
      >< BspAI
      >< Bsp143I
      >< BsiBI
      >< BsaBI
      >< BglII
      >< BspWI
ATGTGGCTAT CACAAGGGCA AAAATTGGCA TTTTGTGCAT AATGTCTGAT AGAGATCTTT ATGACAAACT
17860      17870      17880      17890      17900      17910      17920
      >< XbaI
      >< RmaI
      >< MaeI
      >< MaeII
      >< MaeIII
      >< BsrI
GCAATTTACA AGTCTAGAAA TACCACGTCG CAATGTGGCT ACATTACAAG CAGAAAATGT AACTGGACTT
17930      17940      17950      17960      17970      17980      17990
      >< Sau3AI
      >< NdeII
      >< MboII
      >< MboI
      >< FokI
      >< DpnII
      >< DpnI
      >< BspAI
      >< Bsp143I
      >< NlaIV
      >< Eco64I
      >< BscBI
      >< BanI
      >< AccBII
      >< MnlI
      >< DdeI
      >< Tru9I
      >< MseI>< SfcI
      >< BbsI >< BsrI

```

FIGURE 13. 41

```

TTAAGGACT GTAGTAAGAT CATTACTGGT CTTTCATCCTA CACAGGCACC TACACACCTC AGCGTTGATA
18000      18010      18020      18030      18040      18050      18060

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Eco57I
      >< Ecl136I
      >< DsaV
      >< BstOI
      >< BstNI
      >< HindII>< BsiLI
      >< HincII>< ApyI
      >< PleI
      >< NlaIII
      HinfI ><
      AccI ><
TAAAGTTCAA GACTGAAGGA TTATGTGTTG ACATACCAGG CATACCAAAG GACATGACCT ACCGTAGACT
18070      18080      18090      18100      18110      18120      18130

      >< MaeIII
      >< EcoO65I
      >< Eco91I
      >< BstXI
      >< BstPI
      >< BstEII
      >< MaeIII
      >< EcoO65I
      >< Eco91I
      >< BstXI
      >< BstPI
      >< BstEII
      >< HphI
      >< AccII
      >< AciI
      >< Bsp50I
CATCTCTATG ATGGGTTTCA AAATGAATTA CCAAGTCAAT GGTACCCTA ATATGTTTAT CACCCGCGAA
18140      18150      18160      18170      18180      18190      18200

      >< XmnI
      >< MboII
      >< MaeIII
      >< Asp700I
      >< AluI
      >< MaeII
      >< MnlI
      >< SfaNI
      >< RmaI
      >< NlaIII
      >< MaeI
GAAGCTATTC GTCACGTTTC TCGTGGATT GGCTTTGATG TAGAGGGCTG TCATGCAACT AGAGATGCTG
18210      18220      18230      18240      18250      18260      18270

      >< Tru9I
      >< MseI
      >< HpaI
      >< RsaI
      >< GsuI
      >< Csp6I
      >< BpmI
      >< AfaI
      >< RmaI
      >< MnlI
      >< MaeI
      >< AluI
      >< SfcI
      >< BfrI
      >< AfaI
      >< DdeI
      >< AluI
      >< BsrI
TGGGTACTAA CCTACCTCTC CAGCTAGGAT TTTCTACAGG TGTTAACTTA GTAGCTGTAC CGACTGGTTA
18280      18290      18300      18310      18320      18330      18340

      >< ScrFI
      >< MvaI
      >< MnlI
      >< MaeIII
      >< EcoRII
      >< EcoO65I
      >< EcoNI
      >< Eco91I
      >< Ecl136I
      >< DsaV
      >< Tru9I
      >< DraIII
      >< BstPI
      >< BstOI
      >< BstNI
      >< PmeI
      >< BstEII
      >< BsiI
      >< MseI
      >< BsiYI
      >< HphI
      >< BsiLI
      >< DraI
      >< ApyI
      >< BsrI
      >< HindII
      >< HincII
      >< HphI
      >< EcoRI
      >< Tru9I
      >< MseI

```

FIGURE 13.42

```

TGTTGACACT GAAAATAACA CAGAATTCAC CAGAGTTAAT GCAAAACCTC CACCAGGTGA CCAGTTTAAA
18350      18360      18370      18380      18390      18400      18410

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< DsaV
      >< BstOI
      >< BstNI
      >< BsiLI
      >< BsaJI
      >< NlaIII
      >< ApyI
      >< RsaI
      >< DdeI ><
      >< Tru9I>< Csp6I
      >< MseI >< AfaI
CATCTTATAC CACTCATGTA TAAAGGCTTG CCCTGGAATG TAGTGCGTAT TAAGATAGTA CAAATGCTCA
18420      18430      18440      18450      18460      18470      18480

      >< NlaIII
      >< HinPII
      >< Tth111I
      >< Hin6I
      >< HinfI
      >< HhaI
      >< AspI >< PleI >< CfoI >< AluI
GTGATACACT GAAAGGATTG TCAGACAGAG TCGTGTTTCGT CCTTTGGGCG CATGGCTTTG AGCTTACATC
18490      18500      18510      18520      18530      18540      18550

      >< SinI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< Eco47I
      >< Cfr13I
      >< ScaI
      >< BsiZI
      >< RsaI
      >< Bme18I
      >< Csp6I
      >< AvaII
      >< MaeII
      >< AfaI
      >< AsuI
      >< AflIII
      >< MaeIII>< MaeII
AATGAAGTAC TTTGTCAAGA TTGGACCTGA AAGAACGTGT TGTCTGTGTG ACAAACGTGC AACTTGCTTT
18560      18570      18580      18590      18600      18610      18620

      >< TfiI
      >< Tth111I
      >< HinfI
      >< AspI
TCTACTTCAT CAGATACTTA TGCCTGCTGG AATCATTCTG TGGGTTTTGA CTATGTCTAT AACCCATTTA
18630      18640      18650      18660      18670      18680      18690

      >< ScrFI
      >< RsaI ><
      >< MvaI
      >< EcoRII
      >< Ecl136I ><
      >< DsaV
      >< Csp6I ><
      >< BstXI ><
      >< MaeIII
      >< EcoO65I
      >< Eco91I
      >< BstPI
      >< Eco57I> < BstEII
      >< MaeIII >< NlaIII
      >< AfaI ><
TGATTGATGT TCAGCAGTGG GGCTTTACGG GTAACCTTCA GAGTAACCAT GACCAACATT GCCAGGTACA
18700      18710      18720      18730      18740      18750      18760

      >< SfaNI
      >< RmaI
      >< NspI
      >< NspHI

```

FIGURE 13.43

57/83

```

>< NlaIII >< RmaI
>< MaeI >< NlaIII Tru9I ><
>< NlaIII >< BspWI >< MaeI >< NlaIII
> < AflIII >< BspHI MseI ><
TGGAAATGCA CATGTGGCTA GTTGTGATGC TATCATGACT AGATGTTTAG CAGTCCATGA GTGCTTTGTT
18770 18780 18790 18800 18810 18820 18830

>< ThaI
>< MvnI
>< HinPII
>< Hin6I
>< HhaI
>< CfoI
>< BstUI >< EcoNI> < MnlI
>< Bsp50I >< BslI >< Tru9I
>< AccII >< BsiYI >< DdeI >< MseI
AAGCGCGTTG ATTGGTCTGT TGAATACCCT ATTATAGGAG ATGAACTGAG GGTAAATTCT GCTTGCAGAA
18840 18850 18860 18870 18880 18890 18900

>< RsaI
>< Csp6I >< MboII > < NlaIII
>< AfaI >< NlaIII >< BspWI >< BsrI >< BspHI
AAGTACAACA CATGGTTGTG AAGTCTGCAT TGCTTGCTGA TAAGTTTCCA GTTCTTCATG ACATTGGAAA
18910 18920 18930 18940 18950 18960 18970

>< SauI
>< MstII
>< Eco81I
>< DdeI NlaIII ><
>< CvnI >< EspI
>< Bsu36I >< Eco57I MaeIII ><
>< Bse21I >< DdeI
>< AxyI >< CelII
>< AocI >< MnlI >< SfaNI >< Bpu1102I
TCCAAAGGCT ATCAAGTGTG TGCCTCAGGC TGAAGTAGAA TGGAAGTTCT ACGATGCTCA GCCATGTAGT
18980 18990 19000 19010 19020 19030 19040

>< MnlI >< Ksp632I
>< HindIII >< EarI
>< AluI >< MboII >< Eam1104I
GACAAAGCTT ACAAATAGA GGAACCTCTT TATTCTTATG CTACACATCA CGATAAATTC ACTGATGGTG
19050 19060 19070 19080 19090 19100 19110

>< Sau3AI
>< NdeII
>< MboI
>< MaeII> < MaeIII
>< DpnII
>< DpnI
>< BspAI
>< MaeIII >< Bsp143I >< MunI HinfI >
TTTGTGTTGTT TTGGAATTGT AACGTTGATC GTTACCCAGC CAATGCAATT GTGTGTAGGT TTGACACAAG
19120 19130 19140 19150 19160 19170 19180

Zsp2I ><
>< SphI
> < Ppu10I
>< PaeI
>< NspI
>< NspHI
>< NlaIII
>< EcoRII Mph1103I ><

```

FIGURE 1344

58/83

```

                >< Ecl136I
                >< DsaV
                >< BstOI
                >< BstNI
                >< BsiLI
                >< ApyI
                >< PleI
AGTCTTGTCA AACTTGAAC TACCAGGCTG TGATGGTGGT AGTTTGTATG TGAATAAGCA TGCATTCCAC
19190      19200      19210      19220      19230      19240      19250

                >< Tru9I
                > < MunI
                >< TthHB8I
                >< BcgI/a >< TaqI
                >< AluI
                >< MseI
                >< DraI
                >< BcgI
ACTCCAGCTT TCGATAAAAG TGCATTTACT AATTTAAAGC AATTGCCTTT CTTTTACTAT TCTGATAGTC
19260      19270      19280      19290      19300      19310      19320

                >< PleI
                >< NlaIII
                >< BsmAI
                >< HinfI>< Alw26I
                SfaNI ><
                >< MaeII
                BsaAI ><
                AflIII ><
CTTGTGAGTC TCATGGCAAA CAAGTAGTGT CGGATATTGA TTATGTTCCA CTCAAATCTG CTACGTGTAT
19330      19340      19350      19360      19370      19380      19390

                Zsp2I >
                >< ScaI
                Ppu10I ><
                >< RsaINsiI >
                Mph1103I >
                >< SfaNIEcoT22I >
                > < RsaI >< Csp6I
                >< Csp6I
                >< NlaIII> < AfaI >< AfaI
                >< AvaIII>><
TACACGATGC AATTTAGGTG GTGCTGTTTG CAGACACCAT GCAAATGAGT ACCGACAGTA CTTGGATGCA
19400      19410      19420      19430      19440      19450      19460

                >< FokI
TATAATATGA TGATTTCTGC TGGATTTAGC CTATGGATTT ACAAACAATT TGATACTTAT AACCTGTGGA
19470      19480      19490      19500      19510      19520      19530

                >< ScrFI
                >< MvaI
                >< MaeIII
                >< EcoRII
                >< Ecl136I
                >< DsaV
                >< BstOI
                >< BstNI
                >< BsiLI
                >< ApyI
                >< Tru9I
                >< MseI
ATACATTTAC CAGGTTACAG AGTTTAGAAA ATGTGGCTTA TAATGTTGTT AATAAAGGAC ACTTTGATGG
19540      19550      19560      19570      19580      19590      19600

                >< SgrAI
                >< NaeI
                >< MspI
                >< HpaII
                >< HapII
                >< Cfr10I
                >< BspWI
                > < VspI
                > < Tru9I
                > < MseI
                > < AsnI
                > < AseI
ACACGCCGGC GAAGCACCTG TTTCCATCAT TAATAATGCT GTTTACACAA AGGTAGATGG TATTGATGTG
19610      19620      19630      19640      19650      19660      19670

```

FIGURE 13. 45

59/83

```

>< XhoII
>< Sau3AI
>< NdeII
>< MflI
>< MboI
>< DpnII
  >< DpnI
>< BstYI
>< BspAI
  >< Bsp143I
  >< BglII
GAGATCTTTG AAAATAAGAC AACACTTCCT GTTAATGTTG CATTTGAGCT TTGGGCTAAG CGTAACATTA
19680      19690      19700      19710      19720      19730      19740

                                >< MaeIII
                                >< EspI
                                >< DdeI Tru9I ><
                                >< CeliIMseI ><
                                >< Bpu1102I
                                >< Tru9I
                                >< MseI
                                >< AluI
                                >< Fnu4HI
                                >< EcoRV
                                >< Eco32I
  >< BsrI      >< MseI      >< BbvI
AACCAGTGCC AGAGATTAAG ATACTCAATA ATTTGGGTGT TGATATCGCT GCTAATACTG TAATCTGGGA
19750      19760      19770      19780      19790      19800      19810

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< BsgI
                                >< AflIII
CTACAAAAGA GAAGCCCCAG CACATGTATC TACAATAGGT GTCTGCACAA TGACTGACAT TGCCAAGAAA
19820      19830      19840      19850      19860      19870      19880

  >< DdeI>< MboII
CCTACTGAGA GTGCTTGTTT TTTCACTTACT GTCTTGTTTG ATGGTAGAGT GGAAGGACAG GTAGACCTTT
19890      19900      19910      19920      19930      19940      19950

                                SinI ><
                                Sau96I ><
                                NspIV ><
                                NspHII ><
                                NlaIV ><
                                Eco47I ><
                                Cfr13I ><
                                >< BslI
                                BsiZI ><
                                >< BsiYI
                                BscBI ><
                                Bmel8I ><
                                AvaII ><
                                AsuI ><
                                >< Tru9I
                                >< MseI
TTAGAAACGC CCGTAATGGT GTTTTAATAA CAGAAGGTTT AGTCAAAGGT CTAACACCTT CAAAGGGACC
19960      19970      19980      19990      20000      20010      20020

                                >< VspI
                                >< Tru9I
                                >< PleI
                                >< MseI
  >< RmaI
  >< NheI
  >< MaeI
  >< HgaI>< AluI
AGCACAAAGCT AGCGTCAATG GAGTCACATT AATTGGAGAA TCAGTAAAAA CACAGTTTAA CTACTTTAAG
20030      20040      20050      20060      20070      20080      20090

                                >< Tru9I ><
                                >< Tru9I
                                >< MseI ><
                                >< MseI
                                >< AsnI
                                >< TfiI
                                >< HinfI>< AseI
                                >< HinfI
                                >< MseI
                                >< DdeI
                                >< MnlI
                                >< Tru9I ><
                                >< BsmAI
                                >< DdeI

```

FIGURE 1346

60/83

```

    >< AccI
AAAGTAGACG GCATTATTCA ACAGTTGCCT GAAACCTACT TTA CT CAGAG CAGAGACTTA GAGGATTTTA
    20100      20110      20120      20130      20140      20150      20160

    >< TthHB8I
    >< TaqI
        >< SstI
        >< SduI
        >< SacI
    > < PaeR7I
    > < NspIII
        >< NspII
        >< HgiAI
    > < Eco88I
    > < XhoI>< Eco24I
        >< Ecl136II
    >< XcmI
    >< Sau3AI
    >< NdeII
    >< MboI
    >< DpnII
    >< DpnI
    >< BspAI
    >< Bsp143I
AGCCCAGATC ACAAATGGAA ACTGACTTTC TCGAGCTCGC TATGGATGAA TTCATACAGC GATATAAGCT
    20170      20180      20190      20200      20210      20220      20230

    >< TthHB8I
    >< TaqI
    >< SfuI
    >< NspV
    >< LspI
    >< Csp45I
    >< BstBI
    >< Bsp119I
    >< BsiCI
    >< Bpu14I
    >< AsuII >< BcgI
CGAGGGCTAT GCCTTCGAAC ACATCGTTTA TGGAGATTTC AGTCATGGAC AACTTGGCGG TCTTCATTTA
    20240      20250      20260      20270      20280      20290      20300

    >< HphI
    >< HinPII
    >< Hin6I
    >< EspI > < HhaI >< TfiI
    >< DdeI >< HaeII
    >< CelII >< Eco47III >< Tru9I
    >< Bpu1102I > < CfoI >< HinfI >< MseI
    >< BfrI >< Bsp143II >< MnlI
ATGATAGGCT TAGCCAAGCG CTCACAAGAT TCACCACTTA AATTAGAGGA TTTTATCCCT ATGGACAGCA
    20310      20320      20330      20340      20350      20360      20370

    >< MstI
    >< HinPII
    >< Hin6I
    >< HhaI
    >< FspI
    >< FdiII
    >< CfoI
    >< SfaNI >< AviII
CAGTGAAAAA TTACTTCATA ACAGATGCGC AAACAGGTTT ATCAAAATGT GTGTGTTCTG TGATTGATCT
    20380      20390      20400      20410      20420      20430      20440

    >< TthHB8I

```

FIGURE 13.47



61/83

```

>< Tth111I
>< TaqI
>< AspI > < MaeIII MaeIII ><
TTTACTTGAT GACTTTGTCG AGATAATAAA GTCACAAGAT TTGTCAGTGA TTTCAAAAGT GGTCAAGGTT
20450 20460 20470 20480 20490 20500 20510

>< NspI
>< NspHI
>< NlaIII
>< FokI

>< MunI > < NlaIII >< AflIII
ACAATTGACT ATGCTGAAAT TTCATTCATG CTTTGGTGTA AGGATGGACA TGTTGAAACC TTCTACCCAA
20520 20530 20540 20550 20560 20570 20580

>< SfaNI
>< ScrFI
>< MvaI
>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI >< SfaNI
>< BstNI >< RsaI BspWI ><
>< BsiLI > < Csp6I BsmI >
>< BspWI >< ApyI >< AfaI BscCI ><
AACTACAAGC AAGTCAAGCG TGGCAACCAG GTGTTGCGAT GCCTAACTTG TACAAGATGC AAAGAATGCT
20590 20600 20610 20620 20630 20640 20650

>< Eco57I >< MaeIII >< HphI
TCTTGAAAAG TGTGACCTTC AGAATTATGG TGAAAATGCT GTTATACCAA AAGGAATAAT GATGAATGTC
20660 20670 20680 20690 20700 20710 20720

> < RsaI
>< Csp6I
>< Bst1107I >< Tru9I >< AluI
>< AccI >< MseI > < AfaINlaIII ><
GCAAAGTATA CTCAACTGTG TCAATACTTA AATACACTTA CTTTAGCTGT ACCCTACAAC ATGAGAGTTA
20730 20740 20750 20760 20770 20780 20790

>< ScrFI
>< RsaI
>< MvaI
>< EcoRII >< NspBII
>< Ecl136I >< SduI
> < Csp6I >< NspII
>< BstOI >< PvuII>< HgiAI
>< BstNI >< DdeI
>< BsiLI >< Psp5I>< Bsp1286I
>< ApyI >< AluI >< BmyI
>< DsaV>< AfaI >< Alw21I
TTCAC TTTGG TGCTGGCTCT GATAAAGGAG TTGCACCAGG TACAGCTGTG CTCAGACAAT GGTTGCCAAC
20800 20810 20820 20830 20840 20850 20860

>< XhoII
>< Tru9I
>< Sau3AI
>< NdeII
>< TthHB8I >< MseI
>< MflI
>< MboI
>< MamI
>< DpnII
>< TfiI >< DpnI

```

FIGURE 13. 48

FIGURE 13.49

63/83

```

                                >< Zsp2I
                                >< SphI
                                >< Ppu10I
                                >< PaeI
                                >< NspI
                                >< NspHI
                                >< NsiI
                                >< NlaIII
                                > < NlaIII
                                >< Mph1103I
                                >< EcoT22I
                                > < AvaIII
                                >< MnlI
GGGCTAACTA TCTTGGCAAG CCGAAGGAAC AAATTGATGG CTATACCATG CATGCTAACT ACATTTTCTG
  21220      21230      21240      21250      21260      21270      21280

                                Tru9I ><
                                >< Tru9I
                                MseI ><
                                >< MseI
                                MnlI ><
                                >< MboII
                                >< GsuI
                                >< BsrI
                                >< BpmI
                                >< BbsI
                                >< NlaIII
                                >< MnlI
GAGGAACACA AATCCTATCC AGTTGTCTTC CTATTCACCTC TTTGACATGA GCAAATTTCC TCTTAAATTA
  21290      21300      21310      21320      21330      21340      21350

                                >< Tru9I
                                >< MseI
                                >< Esp4I> < TfiI
                                >< BsmAI
                                >< Alw26I
                                >< AflII> < HinfI
                                Ksp632I ><
                                >< MboII
                                >< EarI
                                Eam1104I ><
AGAGGAACTG CTGTAATGTC TCTTAAGGAG AATCAAATCA ATGATATGAT TTATTCTCTT CTGGAAAAAG
  21360      21370      21380      21390      21400      21410      21420

                                >< Tru9I
                                >< MseI
                                >< HindII
                                >< HincII
                                >< HpaI AflIII >
GTAGGCTTAT CATTAGAGAA AACAACAGAG TTGTGGTTTC AAGTGATATT CTTGTTAACA ACTAAACGAA
  21430      21440      21450      21460      21470      21480      21490

                                >< VneI
                                >< SnoI
                                >< SduI
                                >< NspII
                                >< HpaII
                                >< HgiAI
                                >< HapII
                                >< Cfr10I
                                >< Bsp1286I
                                >< MspI>< BmyI
                                >< ApaLI
                                >< Alw44I
                                >< AgeI >< Alw21I
                                >< SpeI
                                >< RmaI
                                >< MaeI
                                >< MaeIII
                                >< AgeI >< Alw21I
CATGTTTATT TTCTTATTAT TTCTTACTCT CACTAGTGGT AGTGACCTTG ACCGGTGCAC CACTTTTGTG
  21500      21510      21520      21530      21540      21550      21560

                                > < AluI
                                >< MnlI
GATGTTCAAG CTCCTAATTA CACTCAACAT ACTTCATCTA TGAGGGGGGT TTACTATCCT GATGAAATTT
  21570      21580      21590      21600      21610      21620      21630

                                >< Sau3AI

```

FIGURE 13. 50

```

>< NdeII
>< MboI
>< DpnII
  >< DpnI          >< Tru9I
>< BspAI          >< MseI > < MboII
  >< Bsp143I       >< DdeI
TTAGATCAGA CACTCTTTAT TTAAGTCAGG ATTTATTTCT TCCATTTTAT TCTAATGTGA CAGGGTTTCA
21640      21650      21660      21670      21680      21690      21700

  >< VspI
  >< Tru9I
  >< MseI
  >< AsnI          >< Tru9I          >< FokI
  >< AseI >< MaeII >< MseI >< BbvI    > < Fnu4HI
TACTATTAAT CATACGTTTG GCAACCCTGT CATACCTTTT AAGGATGGTA TTTATTTTGC TGCCACAGAG
21710      21720      21730      21740      21750      21760      21770

          >< BslI
          >< DsaI>< BsiYI          >< NlaIII
          >< BsaJI          > < MaeIII
AAATCAAATG TTGTCCGTGG TTGGGTTTTT GGTCTACCA TGAACAACAA GTCACAGTCG GTGATTATTA
21780      21790      21800      21810      21820      21830      21840

          >< NspI
          >< NspHI
          >< NlaIII
  >< Tru9I          >< MaeIII          >< MaeIII
  >< MseI          >< MaeIII          >< MaeIII
  >< HphI          >< MaeIII          >< MaeIII
TTAACAATTC TACTAATGTT GTTATACGAG CATGTAAGTT TGAATTGTGT GACAACCCTT TCTTTGCTGT
21850      21860      21870      21880      21890      21900      21910

          >< StyI          >< Zsp2I
          >< NlaIII          >< Tru9I
          >< NcoI >< RsaI    >< Ppu10I TthHB8I ><
          >< EcoT14I        >< NsiI          >< TaqI
          >< Eco130I        >< MseI          SfaNI ><
          >< DsaI>< Csp6I    >< Mph1103I    RsaI ><
          >< BssT1I          >< TthHB8I >< EcoT22I    Csp6I ><
          >< BsaJI>< AfaI    >< TaqI >< AvaIII    AfaI ><
TTCTAAACCC ATGGGTACAC AGACACATAC TATGATATTC GATAATGCAT TTAATTGCAC TTTTCGAGTAC
21920      21930      21940      21950      21960      21970      21980

          >< Tru9I
          >< MseI
          >< DraI
ATATCTGATG CCTTTTCGCT TGATGTTTCA GAAAAGTCAG GTAATTTTAA ACACCTACGA GAGTTTGTGT
21990      22000      22010      22020      22030      22040      22050

          >< Sau3AI
          >< NdeII
          >< MboI
          >< DpnII
          >< DpnI
          >< BspAI
  >< Tru9I
  >< MseI
  >< DraI          >< SfcI          Bsp143I ><
TTAAAAATAA AGATGGGTTT CTCTATGTTT ATAAGGGCTA TCAACCTATA GATGTAGTTC GTGATCTACC
22060      22070      22080      22090      22100      22110      22120

          >< Tru9I
          >< MseI          > < Tru9I          >< MseI
          >< MseI          > < MseI          >< MnlI
TTCTGGTTTT AACACTTTGA AACCTATTTT TAAGTTGCCT CTTGGTATTA ACATTACAAA TTTTAGAGCC
22130      22140      22150      22160      22170      22180      22190

```

FIGURE 13.51

65/83

```

> < SduI>< SfcI
> < PvuII
> < Psp5I
> < NspII
> < NspBII
> < MaeII > < Fnu4HI
> < Bsp1286I >< PstI      Tru9I >
> < BmyI>< Fnu4HI      MseI >
>< HphI      >< BbvI      >< AluI      >< BbvI
ATTCTTACAG CCTTTTCACC TGCTCAAGAC ATTTGGGGCA CGTCAGCTGC AGCCTATTTT GTTGGCTATT
22200      22210      22220      22230      22240      22250      22260

>< SfaNI
>< RsaI
> < Csp6I
>< DraI      >< AfaI      >< AlwNI
TAAAGCCAAC TACATTTATG CTCAAGTATG ATGAAAATGG TACAATCACA GATGCTGTTG ATTGTTCTCA
22270      22280      22290      22300      22310      22320      22330

> < Tru9I
> < MseI
>< AluI
AAATCCACTT GCTGAACTCA AATGCTCTGT TAAGAGCTTT GAGATTGACA AAGGAATTTA CCAGACCTCT
22340      22350      22360      22370      22380      22390      22400

>< SauI
>< MstII
>< Eco81I
>< DdeI
>< CvnI
>< Bsu36I
>< Bse21I
>< AxyI      >< TfiI
>< MnlI      >< AocI      >< MnlI      >< HinfI      >< SspI      >< MnlI
AATTTTCAGGG TTGTTCCCTC AGGAGATGTT GTGAGATTCC CTAATATTAC AAAGTTGTGT CTTTGGGAG
22410      22420      22430      22440      22450      22460      22470

>< Zsp2I
>< Ppu10I
>< NsiI
> < NlaIII
>< Mph1103I
>< EcoT22I
>< AvaIII
AGGTTTTTAA TGCTACTAAA TTCCCTTCTG TCTATGCATG GGAGAGAAAA AAAATTTCTA ATTGTGTTGC
22480      22490      22500      22510      22520      22530      22540

>< SduI
>< NspII
>< HgiAI
>< Bsp1286I
>< BmyI      >< Tru9I
>< Alw21I      >< MseI      DdeI ><
TGATTACTCT GTGCTCTACA ACTCAACATT TTTTCAACC TTTAAGTGCT ATGGCGTTTC TGCCACTAAG
22550      22560      22570      22580      22590      22600      22610

>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI

```

66/83

```

>< BspAI
>< Bsp143I
TTGAATGATC TTTGCTTCTC CAATGTCTAT GCAGATTCTT TTGTAGTCAA GGGAGATGAT GTAAGACAAA
22620      22630      22640      22650      22660      22670      22680

>< TfiI
>< HinfI

>< ScrFI
>< MvaI
>< HinfII
>< HinfI
>< HhaI
>< HaeII
>< EcoRII
>< Ecl136I
>< DsaV
>< CfoI
>< BstOI
>< BstNI
>< Bsp143II
>< BsiLI
>< ApyI      > < BsrI
TAGCGCCAGG ACAAACTGGT GTTATTGCTG ATTATAATTA TAAATTGCCA GATGATTTCa TGGGTTGTGT
22690      22700      22710      22720      22730      22740      22750

>< SfaNI
>< RmaI
>< MaeI
CCTTGCTTGG AATACTAGGA ACATTGATGC TACTTCAACT GGTAATTATA ATTATAAATA TAGGTATCTT
22760      22770      22780      22790      22800      22810      22820

>< BsrI
DdeI ><
BfrI ><

>< Sau96I
>< Pali
>< NspIV
> < HindIII
>< HaeIII
>< EcoO109I
>< DraII
>< DdeI
>< Cfr13I
>< BsuRI
>< BsiZI
>< BshI
>< BfrI >< PssI
>< NlaIII >< AsuI>< BsmAI
>< AluI      >< Alw26I
AGACATGGCA AGCTTAGGCC CTTTGAGAGA GACATATCTA ATGTGCCTTT CTCCCCTGAT GGCAAACCTT
22830      22840      22850      22860      22870      22880      22890

>< Tru9I
>< Pali
>< MscI
>< HaeIII
>< EaeI>< MseI
>< Tru9I      >< BsuRI
>< MseI      >< BshI
>< BspMI      >< BalI
GCACCCACCC TGCTCTTAAT TGTTATTGGC CATTAAATGA TTATGGTTTT TACACCACTA CTGGCATTGG
22900      22910      22920      22930      22940      22950      22960

BsrI ><

>< Sau96I ><
>< PalINspIV ><
> < MspI      NspHII ><
>< HaeIII

```

FIGURE 13.53

67/83

```

> < HpaII Eco47I ><
    >< DsaI
> < HapII Cfr13I ><
    >< BsuRISinI ><
    >< GdiII Bsi2I ><
    >< BsaJI
    >< Tru9I >< EaeI Bme18I ><
    >< MseI >< Cfr10I AvaII ><
    >< DraI >< BshI AsuI ><
    >< ScaI
    >< RsaI
    >< Csp6I
    >< AfaI
CTACCAACCT TACAGAGTTG TAGTACTTTC TTTTGAACCT TTAAATGCAC CGGCCACGGT TTGTGGACCA
  22970      22980      22990      23000      23010      23020      23030

    >< Tru9I
    >< Tru9I
    >< PleI
    >< MseI
    >< RsaI
    >< Csp6I
    >< BsrI ><
    > < Tru9I
    > < MseI>< BsrI
    >< MseI >< HinfI >< AfaI
AAATTATCCA CTGACCTTAT TAAGAACCAG TGTGTCAATT TTAATTTTAA TGGACTCACT GGTACTGGTG
  23040      23050      23060      23070      23080      23090      23100

    >< Tru9I
    >< MseI
    >< MboII
    >< HpaI
    >< HindII
    >< HincII
    >< Pali
    >< HaeIII
    >< GdiII
    >< EaeI
    >< BsuRI
    >< BshI
    Tfil ><
    HinfI ><
TGTTAACTCC TTCTTCAAAG AGATTTCAAC CATTTCACA ATTTGGCCGT GATGTTTCTG ATTTCACTGA
  23110      23120      23130      23140      23150      23160      23170

    > < XhoII
    >< TthHB8I
    >< TaqI
    > < Sau3AI
    > < NdeII
    > < MflI
    > < MboI
    > < DpnII
    >< DpnI
    > < BstYI
    > < BspAI
    > < SspI
    >< AlwI >< Bsp143I
    >< HphI
TTCCGTTCTGA GATCCTAAAA CATCTGAAAT ATTAGACATT TCACCTTGCT CTTTGGGGG TGTAAGTGTA
  23180      23190      23200      23210      23220      23230      23240

    >< ScrFI
    >< MvaI
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< BstOI
    >< BstNI
    >< BsiLI
    >< ApyI
    >< Tru9I
    >< MseI
    >< HpaI
    >< HindII
    >< Eco57I
    >< BsgI >< HincII
ATTACACCTG GAACAAATGC TTCATCTGAA GTTGCTGTTT TATATCAAGA TGTTAACTGC ACTGATGTTT
  23250      23260      23270      23280      23290      23300      23310

    >< Sau3AI
    >< NlaIII
    >< NdeII
    >< MboI
    >< DpnII
    >< DpnI
    >< HinfII

```

FIGURE 13 54

```

    >< BspWI
    >< BspAI
    >< SfcI
    CTACAGCAAT TCATGCAGAT CAACTCACAC CAGCTTGGCG CATATATTCT ACTGGAAACA ATGTATTCCA
    23320 23330 23340 23350 23360 23370 23380

    >< Hin6I
    > < HhaI
    >< Bsp143I
    >< AluI> < CfoI
    >< BsrI
    >< TthHB8I
    >< TaqI
    >< Sali
    >< RtrI
    >< NspI
    >< EspI >< NspHI
    >< DdeI >< NlaIII
    >< CelII >< HindII
    >< Bpu1102I>< HincII
    >< HinfI
    GACTCAAGCA GGCTGTCTTA TAGGAGCTGA GCATGTCGAC ACTTCTTATG AGTGCGACAT TCCTATTGGA
    23390 23400 23410 23420 23430 23440 23450

    > < SnaBI
    >< ScaI
    >< RsaI
    >< RmaI
    >< MaeII >< MaeI
    > < Eco105I
    >< Csp6I
    >< RmaI
    >< MaeIII
    > < BsaAI
    >< AluI >< MaeI
    >< AfaI
    GCTGGCATT TGTCTAGTTA CCATACAGTT TCTTTATTAC GTAGTACTAG CCAAAAATCT ATTGTGGCTT
    23460 23470 23480 23490 23500 23510 23520

    >< MunI
    ATACTATGTC TTTAGGTGCT GATAGTTCAA TTGCTTACTC TAATAACACC ATTGCTATAC CTACTAACTT
    23530 23540 23550 23560 23570 23580 23590

    RsaI ><
    >< MnlI
    Csp6I ><
    AfaI ><
    >< SfcI
    TTCAATTAGC ATTACTACAG AAGTAATGCC TGTTTCTATG GCTAAAACCT CCGTAGATTG TAATATGTAC
    23600 23610 23620 23630 23640 23650 23660

    > < TfiI
    > < HinfI
    >< AciI
    ATCTGCGGAG ATTCTACTGA ATGTGCTAAT TTGCTTCTCC AATATGGTAG CTTTTCACACA CAACTAAATC
    23670 23680 23690 23700 23710 23720 23730

    >< VneI
    >< SduI
    >< NspII
    >< HgiAI
    >< SnoI>< DdeI
    >< Bsp1286I
    >< BmyI
    >< BbvI
    >< ApaLI
    >< Alw44I
    >< Alw21I
    >< Fnu4HI
    >< BspAI
    >< AflIII
    >< PmlI
    >< PmaCI
    >< MaeII
    >< Eco72I
    >< BsaAI
    >< BbrPI
    >< DpnII >< AlwI
    >< BspAI
    >< AflIII
    GTGCACTCTC AGGTATTGCT GCTGAACAGG ATCGCAACAC ACGTGAAGTG TTCGCTCAAG TCAAACAAAT
    23740 23750 23760 23770 23780 23790 23800

```

FIGURE 13.55



69/83

```

>< RsaI
>< Csp6I
>< AfaI
GTACAAAACC CCAACTTTGA AATATTTTGG TGGTTTAAAT TTTTCACAAA TATTACCTGA CCCTCTAAAG
23810 23820 23830 23840 23850 23860 23870

>< MnlI
>< MnlI
>< DdeI >< MnlI
CCAATAAGA GGTCTTTTAT TGAGGACTTG CTCTTTAATA AGGTGACACT CGCTGATGCT GGCTTCATGA
23880 23890 23900 23910 23920 23930 23940

>< XhoII
>< Sau3AI
>< StyI
>< RmaI
>< MaeI
>< EcoT14I
>< Eco130I
>< BssT1I
>< BsmI
>< BscCI
>< BsaJI
>< BlnI
>< AvrII
>< RmaI
>< NdeII
>< MflI
>< MboI
>< MaeI
>< VspI
>< HphI
>< DpnII
>< HphI
>< DpnI
>< BstYI
>< BspAI
>< Bsp143I
>< BglII
>< MstI
>< HinPII
>< Hin6I
>< HhaI
>< FspI
>< FdiII
>< CfoI
>< AviII
AGCAATATGG CGAATGCCTA GGTGATATTA ATGCTAGAGA TCTCATTTGT GCGCAGAAGT TCAATGGACT
23950 23960 23970 23980 23990 24000 24010

>< RmaIRsaI ><
>< MnlI >< Fnu4HI >< Fnu4HI Csp6I ><
>< BspWI >< BbvI >< BbvI >< BspWI >< MaeIAfaI ><
TACAGTGTTC CCACCTCTGC TCACTGATGA TATGATTGCT GCCTACACTG CTGCTCTAGT TAGTGGTACT
24020 24030 24040 24050 24060 24070 24080

>< MboII
>< HinPII
>< Hin6I
>< HhaI
>< HaeII
>< Fnu4HI >< Ksp632I
>< CfoI >< EarI
>< FokI >< BspWI >< Eam1104I
>< BbvI >< Bsp143II
GCCACTGCTG GATGGACATT TGGTGCTGGC GCTGCTCTTC AAATACCTTT TGCTATGCAA ATGGCATATA
24090 24100 24110 24120 24130 24140 24150

Tru9I ><
MseI ><
>< MaeIII
GGTTCAATGG CATTGGAGTT ACCCAAATG TTCTCTATGA GAACCAAAAA CAAATCGCCA ACCAATTTAA
24160 24170 24180 24190 24200 24210 24220

MaeII ><
>< TfiI
>< HinfI
>< BbvI
>< AluI
CAAGGCGATT AGTCAAATTC AAGAATCACT TACAACAACA TCAACTGCAT TGGGCAAGCT GCAAGACGTT
24230 24240 24250 24260 24270 24280 24290

>< Tru9I
>< MseI
>< HpaI
>< HindII >< BsmI >< Tru9I >< Tru9I >< BfrI
>< HincII>< BscCI >< MseI >< MseI >< AluI

```

FIGURE 13. 56

70/83

```

GTTAACCAGA ATGCTCAAGC ATTAAACACA CTTGTAAAC AACTTAGCTC TAATTTTGGT GCAATTTCAA
24300      24310      24320      24330      24340      24350      24360

      >< ThaI
      >< SpoI
      >< NruI
      >< MvnI
      >< BstUI      >< TthHB8I
      >< Bsp68I      >< TaqI      >< RsaI
      >< EcoRV      >< Bsp50I      >< MnlI      >< Csp6I      >< Tru9I
      >< Eco32I      >< AccII      >< MnlI      >< AciI      >< AfaI      >< MseI
GTGTGCTAAA TGATATCCTT TCGCGACTTG ATAAAGTCGA GGCGGAGGTA CAAATTGACA GGTTAATTAC
24370      24380      24390      24400      24410      24420      24430

      >< MaeIII      >< BbvI      >< Fnu4HI      BbvI ><
AGGCAGACTT CAAAGCCTTC AAACCTATGT AACACAACAA CTAATCAGGG CTGCTGAAAT CAGGGCTTCT
24440      24450      24460      24470      24480      24490      24500

      >< Fnu4HI
      >< BspWI      >< DdeI      >< HindII
      >< BspWI      >< DdeI      >< HincII
GCTAATCTTG CTGCTACTAA AATGTCTGAG TGTGTTCTTG GACAATCAAA AAGAGTTGAC TTTTGTGGAA
24510      24520      24530      24540      24550      24560      24570

      > < NspI
      > < NspHI
      > < NlaIII
      >< MaeIII
      >< NlaIII      >< MaeII
      >< MboII      >< FokI
      >< Fnu4HI      >< BbsI      BsaAI ><
      >< AciI      >< BbvI      >< AflIII
AGGGCTACCA CCTTATGTCC TTCCACAAAG CAGCCCCGCA TGGTGTGTC TTCCTACATG TCACGTATGT
24580      24590      24600      24610      24620      24630      24640

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< BstOI
      >< BstNI      >< HinPII
      >< MnlI      >< BslI      >< Hin6I
      >< DsaV      >< BsiYI      >< HhaI
      >< BsiLI      >< HaeII
      >< BsaJI      >< HphI      >< CfoI      >< NlaIII
      >< ApyI      >< Bsp143II      >< BspHI      EcoNI ><
GCCATCCCAG GAGAGGAACT TCACCACAGC GCCAGCAATT TGTCATGAAG GCAAAGCATA CTTCCCTCGT
24650      24660      24670      24680      24690      24700      24710

      >< MnlI
      >< BslI      >< Tru9I
      >< BsiYI      >< MseI      >< MnlI
GAAGGTGTTT TTGTGTTTAA TGGCACTTCT TGGTTTATTA CACAGAGGAA CTTCTTTTCT CCACAAATAA
24720      24730      24740      24750      24760      24770      24780

      >< DdeI      >< Tru9I
      >< BsmAI      >< SfaNI
      >< SfcI      >< Alw26I      >< MseI      >< AlwI ><
TACTACAGA CAATACATTT GTCTCAGGAA ATTGTGATGT CGTTATTGGC ATCATTAACA ACACAGTTTA
24790      24800      24810      24820      24830      24840      24850

      >< Sau3AI
      >< NdeII

```

FIGURE 13.57

71/83

```

>< MboI          >< P1eI          > < ScaI
>< DpnII         >< MnlI          > < Ksp632I      > < RsaI
  >< DpnI         >< DdeI   >< HinfI      >< MboII
>< BspAI         >< BspWI         > < Eam1104I      >< Csp6I
  >< Bsp143I      >< AluI          > < EarI   > < AluI   > < AfaI   > < HphI
TGATCCTCTG CAACCTGAGC TTGACTCATT CAAAGAAGAG CTGGACAAGT ACTTCAAAAA TCATACATCA
      24860      24870      24880      24890      24900      24910      24920

  >< Sau3AI
  >< NdeII
  >< MboI
>< MamI
  >< DpnII
    >< DpnI
  >< BspAI
    >< Bsp143I
  >< BsiBI          >< Tru9I          >< HindII
  >< BsaBI          >< MseI          >< HincII      AciI ><
CCAGATGTTG ATCTTGCGCA CATTTCAGGC ATTAACGCTT CTGTCGTCAA CATTCAAAAA GAAATTGACC
      24930      24940      24950      24960      24970      24980      24990

          >< Tru9I
              > < TfiI
  >< MnlI          >< SwaI
>< EcoNI          >< MseI
  >< BslI          > < HinfI
>< MnlI>< BsiYI    >< DraI
GCCTCAATGA GGTGCGTAAA AATTTAAATG AATCACTCAT TGACCTTCAA GAATTGGGAA AATATGAGCA
      25000      25010      25020      25030      25040      25050      25060

  >< StyI
  >< P1eI
  >< HaeIII
  >< EcoT14I
  >< Eco130I
  >< BsuRI
  >< BstT1I
  >< Tru9I>< BshI          NlaIII ><
  >< MseI >< BsaJI          MaeIII ><
          >< BstXI
ATATATTAAA TGGCCTTGGT ATGTTTGGCT CGGCTTCATT GCTGGACTAA TTGCCATCGT CATGGTTACA
      25070      25080      25090      25100      25110      25120      25130

          > < SphI
          > < PaeI
          > < NspI
          > < NspHI
          > < NlaIII
          > < MaeI
          >< MnlI>< BbvI Fnu4HI ><
ATCTTGCTTT GTTGCATGAC TAGTTGTTGC AGTTGCCTCA AGGGTGCATG CTCTGTGGT TCTTGCTGCA
      25140      25150      25160      25170      25180      25190      25200

          >< FokI
          >< DdeI
>< MnlI >< P1eI>< HinfI >< BsrI
AGTTTGATGA GGATGACTCT GAGCCAGTTC TCAAGGGTGT CAAATTACAT TACACATAAA CGAATTATG
      25210      25220      25230      25240      25250      25260      25270

          >< Sau3AI
          >< NdeII
          >< MboI
          >< DpnII
          > < DpnI

```

FIGURE 13.58

72/83

```

>< BspAI
> < Bsp143I
>< BsgI >< AlwI >< BsrI BspWI >
GATTTGTTTA TGAGATTTT TACTCTTGGA TCAATTACTG CACAGCCAGT AAAAATTGAC AATGCTTCTC
25280 25290 25300 25310 25320 25330 25340

>< ScaI
>< RsaI
>< Csp6I >< SfcI
>< AfaI >< NlaIII >< AciI >< MnlI FokI >
CTGCAAGTAC TGTTTCATGCT ACAGCAACGA TACCGCTACA AGCCTCACTC CCTTTCGGAT GGCTTGTTAT
25350 25360 25370 25380 25390 25400 25410

> < HinPII
> < Hin6I
>< HhaI RmaI ><
>< HaeII >< HinPII NheI ><
>< Eco47III >< Hin6I MaeI ><
>< CfoI >< HhaI Fnu4HI ><
>< BspWI >< Bsp143II >< CfoI AluI ><
TGCGGTTGCA TTTCTTGCTG TTTTTCAGAG CGCTACCAA ATAATTGCGC TCAATAAAAG ATGGCAGCTA
25420 25430 25440 25450 25460 25470 25480

>< EcoNI
>< BslI
>< BsiYI >< MaeIII
>< BbvI >< BsrI >< BbvI > < Fnu4HI BbvI ><
GCCCTTTATA AGGGCTTCCA GTTCATTTGC AATTTACTGC TGCTATTTGT TACCATCTAT TCACATCTTT
25490 25500 25510 25520 25530 25540 25550

>< SfcI >< HinPII
>< PstI >< Hin6I >< RsaI NsiI ><
> < Fnu4HI >< HhaI >< Csp6I Mph1103I ><
>< BspMI >< MnlI >< CfoI >< AfaI >< MnlI AvaIII ><
TGCTTGTCGC TGCAGGTATG GAGGCGCAAT TTTTGACCT CTATGCCTTG ATATATTTTC TACAATGCAT
25560 25570 25580 25590 25600 25610 25620

>< SfaNI
>< NspI
>< NspHI
>< NlaIII >< SfaNI
CAACGCATGT AGAATTATTA TGAGATGTTG GCTTTGTTGG AAGTGCAAAT CCAAGAACCC ATTACTTTAT
25630 25640 25650 25660 25670 25680 25690

>< Bst1107I
>< AccI MaeIII ><
GATGCCAACT ACTTTGTTTG CTGGCACACA CATAACTATG ACTACTGTAT ACCATATAAC AGTGTCACAG
25700 25710 25720 25730 25740 25750 25760

>< MboII
>< HphI BstXI ><
>< MunI >< MaeIII >< MaeIII >< Eco57I >< BbsI MnlI >
ATACAATTGT CGTTACTGAA GGTGACGGCA TTTCAACACC AAAACTCAAA GAAGACTACC AAATTGGTGG
25770 25780 25790 25800 25810 25820 25830

>< RsaI
> < NlaIII
>< HphI
>< Tru9I >< Tth111I>< Csp6I
>< DdeI >< DdeI >< MseI>< AspI >< AfaI

```

FIGURE 13.59

73/83

```

TTATTCTGAG GATAGGCACT CAGGTGTTAA AGACTATGTC GTTGATCATG GCTATTTTAC CGAAGTTTAC
25840      25850      25860      25870      25880      25890      25900

                                Tru9I ><
                                MseI ><
    > < HinfI>< PleI          >< BsrI
    >< AluI >< AccI          >< SfcI >< AlwNI          >< MboII          HindIII >
TACCAGCTTG AGTCTACACA AATTACTACA GACACTGGTA TTGAAAATGC TACATTCTTC ATCTTTAACA
25910      25920      25930      25940      25950      25960      25970

                                > < TthHB8I
                                > < TaqI          >< Ksp632I
                                > < MboII          >< EarI BspWI ><
    >< AluI
    >< MseI
    >< Eco57I          >< Eam1104I AlwI ><
AGCTTGTTAA AGACCCACCG AATGTGCAAA TACACACAAT CGACGGCTCT TCAGGAGTTG CTAATCCAGC
25980      25990      26000      26010      26020      26030      26040

    >< XhoII
    >< Sau3AI
    >< NlaIV
    >< NdeII
    >< MflI
    >< MboI
    >< DpnII
    >< DpnI
    >< BstYI
    >< BstI
    >< BspAI
    >< Bsp143I
    >< BscBI
    >< BamHI >< AlwI
AATGGATCCA ATTTATGATG AGCCGACGAC GACTACTAGC GTGCCTTTGT AAGCACAAGA AAGTGAGTAC
26050      26060      26070      26080      26090      26100      26110

                                >< RmaI
                                >< MaeI
                                >< RsaI ><
                                Csp6I ><
                                AfaI ><
                                > < Tru9I
    >< RsaI
    > < MseI
    >< MboII
    >< MaeII
    >< RsaI
    >< Csp6I
    >< AfaI
    >< Csp6I
    >< AfaI
    >< Tru9I >< Csp6I
    >< MseI >< AfaI
GAACTTATGT ACTCATTCGT TTCGGAAGAA ACAGGTACGT TAATAGTTAA TAGCGTACTT CTTTTTCTTG
26120      26130      26140      26150      26160      26170      26180

                                >< TthHB8I
                                >< TaqI
                                >< RmaI
                                >< HinfII
                                > < RsaI
                                > < MaeIII
                                >< HinfI
                                Fnu4HI ><
                                >< MaeI >< RmaI
                                >< HhaI
                                >< Csp6I
                                >< FokI >< MaeI
                                >< CfoI >< BbvI > < AfaI
CTTTCGTGGT ATTCTTGCTA GTCACACTAG CCATCCTTAC TGCGCTTCGA TTGTGTGCGT ACTGCTGCAA
26190      26200      26210      26220      26230      26240      26250

                                >< Tru9I
                                >< ThaI
                                >< MvnI
                                >< MseI
                                >< BstUI
                                Ksp632I >
                                >< MaeII >< Bsp50I >< MboII EarI >
                                >< AccI >< AccII
                                Eam1104I >
TATTGTTAAC GTGAGTTTGT TAAAACCAAC GGTTTACGTC TACTCGCGTG TTAAAAATCT GAACTCTTCT
26260      26270      26280      26290      26300      26310      26320

```

FIGURE 13.60

74/83

```

>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< MboII>< DpnI
>< XmnI >< BspAI> < Eco57I
>< Asp700I>< Bsp143I
GAAGGAGTTC CTGATCTTCT GGTCTAAACG AACTAACTAT TATTATTATT CTGTTTGGAA CTTTAACATT
26330      26340      26350      26360      26370      26380      26390

>< ScrFI
>< MvaI
>< EcoRII
>< Ecl136I
>< DsaV NlaIV ><
>< BstOI
>< BstNI RmaI ><
>< BsiLI MaeI ><
>< ApyIBscBI ><
>< NlaIII >< AfaI >< AluI
GCTTATCATG GCAGACAACG GTACTATTAC CGTTGAGGAG CTTAAACAAC TCCTGGAACA ATGGAACCTA
26400      26410      26420      26430      26440      26450      26460

>< ScrFI
>< RmaI
>< MvaI
>< MaeI
>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< ApyI >< MaeIII
GTAATAGGTT TCCTATTCCT AGCCTGGATT ATGTTACTAC AATTTGCCTA TTCTAATCGG AACAGGTTTT
26470      26480      26490      26500      26510      26520      26530

>< Pali
>< MscI
>< MnlI >< MaeIII
>< HaeIII
>< EaeI
>< BsuRI
>< BsrI
>< BspWI
>< RsaI
>< Csp6I >< HindIII
>< AfaI >< AluI
>< BshI
>< BalI
>< BbvI Fnu4HI ><
TGTACATAAT AAAGCTTGTT TTCCTCTGGC TCTTGTGGCC AGTAACACTT GCTTGTTTTG TGCTTGCTGC
26540      26550      26560      26570      26580      26590      26600

>< VspI
>< Tru9I
>< MseI
>< HphI
>< SfcI >< AsnI >< BsrI
>< AccI >< AseI>< MaeIII>< AciI
TGTCTACAGA ATTAATTGGG TGA CTGGCGG GATTGCGATT GCAATGGCTT GTATTGTAGG CTTGATGTGG
26610      26620      26630      26640      26650      26660      26670

>< EspI
>< Eco57I
>< DdeI
>< CeiII
>< Bpu1102I
>< RsaI
>< Csp6I

```

FIGURE 13.61

75/83

```

>< BfrI
  >< AluI
CTTAGCTACT TCGTTGCTTC CTCAGGCTG TTTGCTCGTA CCCGCTCAAT GTGGTCATTC AACCCAGAAA
  26680      26690      26700      26710      26720      26730      26740

  >< ScrFI
  >< NciI
  >< MspI
  >< HpaII
  >< HapII
  >< DsaV>< MnlI
  >< BslI
  >< BsiYI
  >< BsaJI >< MunI          > < XcmI
  >< BcnI      >< MaeIII >< AciI >< NlaIII
CAAACATTCT TCTCAATGTG CCTCTCCGGG GGACAATTGT GACCAGACCG CTCATGGAAA GTGAACTTGT
  26750      26760      26770      26780      26790      26800      26810

Tru9I ><
  SinI >
  Sau96I >
  PpuMI >
  NspIV >
  MseI ><
  >< MaeIII
  >< Sau3AI          > < RmaI >< HaeII
  >< NdeII          >< Pali > < MaeI EcoO109I >
  >< MboI          >< MspI          >< HinPIIEco47I >
  >< FbaI          >< HpaII >< StyI>< Hin6I DraII >
  >< DpnII          >< HapII >< EcoT14I Cfr13I >
  >< DpnI          >< HaeIII >< Eco130I>< Bsp143II
  >< BspAI          >< GdiII          >< BssTII Bsi2I >
  >< Bsp143I        >< EaeI          >< BsaJI Bml18I >
  >< BsiQI          >< BsuRI          >< BlnI >< HhaI AvaII >
  >< BclI          >< MaeIII          >< BshI >< AvrII >< CfoI AsuI >
CATTGGTGCT GTGATCATTC GTGGTCACTT GCGAATGGCC GGACACTCCC TAGGGCGCTG TGACATTAAG
  26820      26830      26840      26850      26860      26870      26880

  >< Sau3AI
  >< NdeII
  >< MboI
  >< DpnII
  >< DpnI

  >< PssI >< BspMI
  >< Psp5II >< BspAI          >< XmnI
  >< NspHII >< Bsp143I          >< Asp700I > < HgaI Fnu4HI ><
GACCTGCCAA AAGAGATCAC TGTGGCTACA TCACGAACGC TTTCTTATTA CAAATTAGGA GCGTCGCAGC
  26890      26900      26910      26920      26930      26940      26950

  >< TfiI
  >< HinfI
  >< BbvI          > < Tru9I
  >< BbvI          >< Fnu4HI >< AciI          > < MseI
GTGTAGGCAC TGATTCAGGT TTTGCTGCAT ACAACCGCTA CCGTATTGGA AACTATAAAT TAAATACAGA
  26960      26970      26980      26990      27000      27010      27020

  >< MspI          >< RsaI
  >< HpaII          >< RmaI
  >< HapII          >< Csp6I
  >< Cfr10I        >< MaeI>< BcgI          HindII ><
  >< BcgI/a        >< SspI          >< AfaI >< MaeIII      HincII ><

```

FIGURE 13.62

CCACGCCGGT AGCAACGACA ATATTGCTTT GCTAGTACAG TAAGTGACAA CAGATGTTTC ATCTTGTTGA  
 27030 27040 27050 27060 27070 27080 27090

>< ScrFI  
 >< MvaI  
 >< MaeIII  
 >< EcoRII  
 >< Ecl136I  
 >< DsaV  
 >< BstOI  
 >< BstNI  
 >< BsiII  
 >< ApyI

>< MnlI HinfI >< TfiI  
 CTTCCAGGTT ACAATAGCAG AGATATTGAT TATCATTATG AGGACTTTCA GGATTGCTAT TTGGAATCTT  
 27100 27110 27120 27130 27140 27150 27160

>< BsmAI >< Tru9I >< MnlI  
 >< MaeII >< Alw26I >< MseI >< DdeI >< MboII  
 GACGTTATAA TAAGTTCAAT AGTGAGACAA TTATTTAAGC CTCTAACTAA GAAGAATTAT TCGGAGTTAG  
 27170 27180 27190 27200 27210 27220 27230

>< Ksp632I  
 >< MboII >< EarI  
 >< MboII >< NlaIII Eam1104I ><  
 ATGATGAAGA ACCTATGGAG TTAGATTATC CATAAAACGA ACATGAAAAT TATTCTCTTC CTGACATTGA  
 27240 27250 27260 27270 27280 27290 27300

>< RsaI >< RsaI  
 >< Csp6I >< Csp6I  
 >< AluI >< MnlI >< AfaI >< AfaI  
 TTGTATTTAC ATCTTGCGAG CTATATCACT ATCAGGAGTG TGTAGAGGT ACGACTGTAC TACTAAAAGA  
 27310 27320 27330 27340 27350 27360 27370

>< MnlI >< HphI >< HphI >< MnlI  
 ACCTTGCCCA TCAGGAACAT ACGAGGGCAA TTCACCATT CACCCTCTTG CTGACAATAA ATTTGCACTA  
 27380 27390 27400 27410 27420 27430 27440

Sau3AI >  
 >< PvuII  
 >< Psp5I  
 >< NspBII  
 >< TthHB8I NdeII >  
 >< TaqI MboI >  
 >< RsaI >< Fnu4HI  
 >< Csp6I DpnII >  
 >< BbvI BspAI >  
 >< AfaI >< AluI  
 ACTTGCACTA GCACACACTT TGCTTTTGCT TGTGCTGACG GTACTCGACA TACCTATCAG CTGCGTGCAA  
 27450 27460 27470 27480 27490 27500 27510

>< SstI  
 >< SduI  
 >< SacI  
 >< NspII  
 >< HgiAI  
 >< Eco24I  
 >< Ecl136II  
 >< BspWI  
 >< Bsp1286I  
 >< BmyI  
 >< BanII  
 >< Alw21I

>< HphI  
 >< DpnI

>< MnlI



77/83

```

>< Bsp143I          >< MnlI          > < AluI          BbvI ><
GATCAGTTTC ACCAAAAC TTTCATCAGAC AAGAGGAGGT TCAACAAGAG CTCTACTCGC CACTTTTTCT
  27520      27530      27540      27550      27560      27570      27580

                                SstI ><
                                SduI ><
                                SacI ><
                                NspII ><
                                HgiAI ><
                                Eco24I ><
                                Ecl136II ><
                                Bsp1286I ><
                                BmyI ><
                                BanII ><
                                >< RmaI   >< Tru9I   >< Tru9I   Alw21I ><
                                >< MaeI   >< MseI   >< MseI   AluI ><
>< Fnu4HI          >< HphI          >< MseI          >< MseI          >< MseI
CATTGTTGCT GCTCTAGTAT TTTTAATACT TTGCTTCACC ATTAAGAGAA AGACAGAATG AATGAGCTCA
  27590      27600      27610      27620      27630      27640      27650

>< Tru9I
>< MseI
CTTTAATTGA CTTCTATTTG TGCTTTTTAG CCTTTCTGCT ATTCCTTGTT TTAATAATGC TTATTATATT
  27660      27670      27680      27690      27700      27710      27720

                                >< XhoII
                                >< XbaI
> < ScrFI
                                >< Sau3AI
                                >< RmaI
                                >< NdeII
> < MvaI
                                >< MflI
                                >< MboI
>< EcoRII>< MaeI
> < Ecl136I
                                >< DpnII
                                >< DpnI
                                >< BstYI
> < BstOI
> < BstNI
>< TthHB8I >< BspAI          > < RsaI
                                >< DsaV>< Bsp143I   >< MboII
                                > < BsiLI   >< Csp6I
>< TaqI > < ApyI   > < AlwI   > < AfaI   >< NlaIII
TTGGTTTTCA CTCGAAATCC AGGATCTAGA AGAACCTTGT ACCAAAGTCT AAACGAACAT GAAACTTCTC
  27730      27740      27750      27760      27770      27780      27790

                                >< HinPII
                                >< Hin6I
                                >< HhaI
                                >< RsaI >< HaeII
                                >< SfcI   >< Eco47III
                                >< Csp6I>< CfoI   SfaNI ><
                                >< AfaI >< Bsp143II
ATTGTTTTGA CTTGTATTTT TCTATGCAGT TGCATATGCA CTGTAGTACA GCGCTGTGCA TCTAATAAAC
  27800      27810      27820      27830      27840      27850      27860

                                >< XhoII
                                >< Sau3AI
                                >< NdeII
> < MnlI
                                >< MflI

```

FIGURE 13.64

78/83

```

>< MboI
>< DpnII
    >< DpnI      >< RsaI
    >< BstYI    >< MboII
>< NlaIII>< BspAI      >< Csp6I >< RmaI
    >< AlwI >< Bsp143I  >< AfaI >< MaeI
CTCATGTGCT TGAAGATCCT TGTAAGGTAC AACACTAGGG GTAATACTTA TAGCACTGCT TGGCTTTGTG
    27870      27880      27890      27900      27910      27920      27930

>< SduI
>< RmaI
>< NspII
>< MaeI
>< HgiAI
>< Bsp1286I
>< BmyI
>< Alw21I
                                >< NspI
                                >< NspHI
                                >< NlaIII >< MaeIII
CTCTAGGAAA GGTTTTACCT TTTCATAGAT GGCACACTAT GGTTCAAACA TGCACACCTA ATGTTACTAT
    27940      27950      27960      27970      27980      27990      28000

    > < XhoII
    > < Sau3AI > < Van91I
        >< PvuII
        >< Psp5I
    > < NdeII > < PflMI
    > < MflI>< NspBII
    > < DpnII      >< HinP1I
        >< Bsp143I  >< Hin6I
    > < BstYI > < BslI >< HhaI >< RmaI
    > < BspAI > < BsiYI>< CfoI >< MaeI
    > < MboI>< AluI>< BspWI >< BspWI
>< AlwI >< DpnI > < AccB7I >< AluI
CAACTGTCAA GATCCAGCTG GTGGTGCCT TATAGCTAGG TGTTGGTACC TTCATGAAGG TCACCAAAC
    28010      28020      28030      28040      28050      28060      28070

                                >< SinI
                                >< Sau96I
                                >< NspIV
                                NspHII ><
                                NlaIV ><
                                >< Eco47I
                                >< Cfr13I
                                >< BsiZI
                                BscBI ><
                                >< Bme18I
                                >< AvaII
                                >< AsuI
>< Fnu4HI >< MaeII
    >< Esp3I >< Csp6I >< Tru9I
    >< BsmAI >< BsmBI >< MseI
    >< Alw26I >< AfaI >< DraI
GCTGCATTTA GAGACGTACT TGTGTTTTTA AATAAACGAA CAAATTAAAA TGTCTGATAA TGGACCCCAA
    28080      28090      28100      28110      28120      28130      28140

                                >< SinI
                                >< Sau96I
                                >< NspIV
                                >< NspHII
                                >< NlaIV
                                >< Eco47I
                                >< Cfr13I
                                >< BsiZI
                                >< BscBI
                                >< Bme18I
                                >< AvaII
                                >< TfiI
                                >< AsuI
                                >< HinfI
                                >< MnlI

    >< SduI
    >< NspII
    >< Bsp1286I
    >< BmyI
>< MaeII >< AciI

```

FIGURE 13. 65

79/83

```

TCAAACCAAC GTAGTGCCCC CCGCATTACA TTTGGTGGAC CCACAGATTC AACTGACAAT AACCAGAATG
28150      28160      28170      28180      28190      28200      28210

      >< HinPII >< StyI
      >< HaeII
      > < Pali >< Hin6I >< EcoT14I
      > < HaeIII >< HhaI>< Eco130I
      >< BspWI >< BssTII
      > < BsuRI >< Bsp143II
      >< HgaI> < BshI >< CfoI>< BsaJI >< HgaI
GAGGACGCAA TGGGGCAAGG CCAAAACAGC GCCGACCCCA AGGTTTACCC AATAATACTG CGTCTTGGTT
28220      28230      28240      28250      28260      28270      28280

      >< TthHB8I
      > < ScrFI
      >< Pali
      >< PaeR7I
      >< NspIII
      > < MvaI
      >< HaeIII
      >< EcoRII
      >< Eco88I
      >< XhoI > < Ecl136I
      >< DsaV
      >< BsuRI
      >< SlaI > < BstOI
      >< MnlI>< TaqI> < BstNI
      >< CcrI > < BsiLI
      >< HinfI >< BshI
      >< TfiI>< BcoI>< BsaJI
      >< MnlI >< DdeI >< AvaI > < ApyI
      >< AluI >< DdeI > < NlaIII >< BfrI >< Ama87I >< MnlI
CACAGCTCTC ACTCAGCATG GCAAGGAGGA ACTTAGATTC CCTCGAGGCC AGGGCGTTCC AATCAACACC
28290      28300      28310      28320      28330      28340      28350

      >< SniI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< Eco47I
      >< Cfr13I
      >< Bsi2I
      >< Bme18I > < Ksp632I
      >< AvaII > < Eam1104I
      >< AsuI > < EarI > < AluI>< MboII >< MaeIII
AATAGTGGTC CAGATGACCA AATTGGCTAC TACCGAAGAG CTACCCGACG AGTTTCGTGGT GGTGACGGCA
28360      28370      28380      28390      28400      28410      28420

      >< SstI
      >< SduI
      >< SacI
      >< NspII
      >< HgiAI
      >< EspI
      >< Eco24I
      >< Ecl136II >< StyI >< Sau96I
      >< DdeI >< RmaI >< Pali
      >< CelII >< MaeI >< NspIV
      >< Bsp1286I >< EcoT14I >< Cfr13I
      >< Bpu1102I >< Eco130I >< BsuRI
      >< BmyI >< BssTII > < BsrI
      >< BanII >< RsaI >< BsaJI >< Bsi2I

```

FIGURE 13. 66

80/83

```

>< Alw21I    >< Csp6I    >< BlnI    >< BshI>< HindIII
>< HphI  >< AluI    >< AfaI    >< AvrII  >< AsuI    >< AluI
AAATGAAAGA GCTCAGCCCC AGATGGTACT TCTATTACCT AGGAAGCTGC CCAGAAGCTT CACTTCCCTA
28430      28440      28450      28460      28470      28480      28490

>< HinP1I
>< Hin6I
>< HhaI
>< HaeII
>< CfoI
>< Bsp143II
CGGCGCTAAC AAAGAAGGCA TCGTATGGGT TGCAACTGAG GGAGCCTTGA ATACACCCAA AGACCACATT
28500      28510      28520      28530      28540      28550      28560

>< NlaIV
>< Eco64I
>< BscBI
>< BanI
>< AciI
>< AccB1I >< BbvI    >< Fnu4HI    >< MnlI
GGCACCCGCA ATCCTAATAA CAATGCTGCC ACCGTGCTAC AACTTCCTCA AGGAACAACA TTGCCAAAAG
28570      28580      28590      28600      28610      28620      28630

>< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI
>< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI
>< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI    >< MnlI
GCTTCTACGC AGAGGGAAGC AGAGGCGGCA GTCAAGCCTC TTCTCGCTCC TCATCACGTA GTCGCGGTAA
28640      28650      28660      28670      28680      28690      28700

>< ScrFI
>< MvaI
>< EcoRII
>< Ecl136I
>< DsaV>< Fnu4HI
>< BstOI
>< BstNI
>< BsiLI
>< ApyI
>< BbvI
>< TaqI
>< AciI
TTCAAGAAAT TCAACTCCTG GCAGCAGTAG GGGAAATTCT CCTGCTCGAA TGGCTAGCGG AGGTGGTGAA
28710      28720      28730      28740      28750      28760      28770

> < ThaI
> < MvnI
>< HphI >< MnlI
> < HinP1I
> < Hin6I
>< HhaI
> < BstUI >< RmaI
> < Bsp50I >< MaeI
>< BbvI >< CfoI>< Fnu4HI
> < AccII>< BspWI
>< AluI
ACTGCCCTCG CGCTATTGCT GCTAGACAGA TTGAACCAGC TTGAGAGCAA AGTTTCTGGT AAAGGCCAAC
28780      28790      28800      28810      28820      28830      28840

> < PalI>< MaeIII
> < HaeIII
> < BsuRI >< DdeI
>< Fnu4HI
>< DdeI
>< RsaI ><
>< MnlI
>< MaeII ><
>< Csp6I ><

```

FIGURE 13.67

81/83

```

> < BshI > < BbvI > < MnlI > < BspWI > < SfaNI > AfaI >
AACACAAGG CCAAACTGTC ACTAAGAAAT CTGCTGCTGA GGCATCTAAA AAGCCTCGCC AAAAACGTAC
28850      28860      28870      28880      28890      28900      28910

>> Tth111I
>> SinI
>> Sau96I
>> NspIV
>> NspHII
> < MaeII
>> Eco47I
>> Cfr13I
>> BsmBI
>> BsiZI >> StyI
>> Bme18I >> EcoT14I
>> AvaII >> Eco130I
>> AsuI >> BssT1I
>> BsaJI
>> RsaI
>> MaeIII
>> MaeII
>> Csp6I
>> AfaI
>> Esp3I
>> BsmAI
>> Alw26I > < AspI
>> BsaJI
TGCCACAAAA CAGTACAACG TCACTCAAGC ATTTGGGAGA CGTGGTCCAG AACAAACCCA AGGAAATTTTC
28920      28930      28940      28950      28960      28970      28980

>> SinI
>> Sau96I
>> NspIV
>> NspHII
>> NlaIV
>> Eco47I
>> Cfr13I
>> BsiZI
>> BscBI
>> Bme18I
>> AvaII
>> AsuI
>> Pali
>> HaeIII
>> GdiII
>> Fnu4HI
>> EaeI
>> BsuRI
>> BshI
>> AciI
>> BspWI >
>> BspWI
GGGGACCAAG ACCTAATCAG ACAAGGAACT GATTACAAAC ATTGGCCGCA AATTGCACAA TTTGCTCCAA
28990      29000      29010      29020      29030      29040      29050

>> BsmI
>> BscCI >> MnlI >> MaeIII
>> NlaIII
>> MaeIII
>> NlaIII
GTGCCTCTGC ATTCTTTGGA ATGTCACGCA TTGGCATGGA AGTCACACCT TCGGGAACAT GGCTGACTTA
29060      29070      29080      29090      29100      29110      29120

>> XhoII
>> Sau3AI
>> NdeII
>> MflI
>> MboI
>> FokI
>> Tru9I
>> NlaIV
>> NlaIII
>> MseI
>> BscBI >> BstXI >> AlwI > < Bsp143I
>> AspI
>> Tth111I
>> MaeII
>> BspWI >
TCATGGAGCC ATTAAATTGG ATGACAAAGA TCCACAATTC AAAGACAACG TCATACTGCT GAACAAGCAC
29130      29140      29150      29160      29170      29180      29190

>> HgaI
>> EspI >>
>> DdeI >>
>> CelII >>
>> Bpu1102I >>
>> AluI >>
ATTGACGCAT ACAAACATT CCCACCAACA GAGCCTAAAA AGGACAAAAA GAAAAAGACT GATGAAGCTC
29200      29210      29220      29230      29240      29250      29260

```

FIGURE 13.68

```

                                >< PleI
                                >< MboII
                                >< Ksp632I >< GsuI
                                >< MboII
                                >< MaeIII >< EarI>< Fnu4HI
                                >< HinfI >< Eam1104I>< BpmI
                                >< Fnu4HI >< BbvI >< AciI >< NlaIII
AGCCTTTGCC GCAGAGACAA AAGAAGCAGC CCACTGTGAC TCTTCTTCCT GCGGCTGACA TGGATGATTT
29270      29280      29290      29300      29310      29320      29330

                                >< NlaIII >< HinfI NlaIII ><
                                >< AluI >< TfiI>< DdeI >< BspHI
CTCCAGACAA CTTCAAAATT CCATGAGTGG AGCTTCTGCT GATTCAACTC AGGCATAAAC ACTCATGATG
29340      29350      29360      29370      29380      29390      29400

                                >< MaeII >< AccI
ACCACACAAG GCAGATGGGC TATGTAAACG TTTTCGCAAT TCCGTTTACG ATACATAGTC TACTCTTGTG
29410      29420      29430      29440      29450      29460      29470

                                >< Tru9I
                                >< Tru9I
                                >< MseI
                                >< MseI
                                >< HpaI
                                >< HindII Tru9I ><
                                >< HincII MseI ><
CAGAATGAAT TCTCGTAACT AAACAGCACA AGTAGGTTTA GTTAACTTTA ATCTCACATA GCAATCTTTA
29480      29490      29500      29510      29520      29530      29540

                                XorII >
                                TthHB8I >
                                TaqI >
                                Sau3AI ><
                                RsaI ><
                                >< ThaIPvuI >
                                NdeII ><
                                >< MnlI
                                >< MvnIMcrI >
                                MboI ><
                                DpnII ><
                                DpnI ><
                                Csp6I ><
                                >< BstUI
                                >< HaeIII BspCI >
                                BspAI ><
                                >< TthHB8I >< Bsp50I
                                >< Pali Bsp143I ><
                                >< BsuRI BsiEI >
                                >< BshIAfaI ><

                                >< MnlI
                                >< MaeIII
ATCAATGTGT AACATTAGGG AGGACTTGAA AGAGCCACCA CATTTTCATC GAGGCCACGC GGAGTACGAT
29550      29560      29570      29580      29590      29600      29610

                                >< TaqI >< AciI
                                >< MnlI >< AccII

                                >< SduI
                                >< NspII
                                >< MboII >< VspI
                                >< Ksp632I >< Eco24I >< Tru9I
                                >< Bsp1286I >< MseI
                                >< EarI >< BmyI >< AsnI
                                >< AfaI >< BbvI >< AluI>< Eam1104I >< BanII >< AseI

```

FIGURE 13.69

83/83

CGAGGGTACA GTGAATAATG CTAGGGAGAG CTGCCTATAT GGAAGAGCCC TAATGTGTAA AATTAATTTT  
29620 29630 29640 29650 29660 29670 29680

>< Tru9I >< DdeI

>< MseI >< BfrI

>< NlaIII > < AluI

AGTAGTGCTA TCCCCATGTG ATTTTAATAG CTTCTTAGGA GAATGACAAA AAAAAAAAAA AAAAAA  
29690 29700 29710 29720 29730 29740

S226CAS111.ST25  
SEQUENCE LISTING

<110> INSTITUT PASTEUR  
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

<120> utilisation des protéines et des peptides codés par le génome d'une nouvelle souche de coronavirus associé au SRAS

<130> F226CAS111

<160> 75

<170> PatentIn version 3.1

<210> 1

<211> 29746

<212> DNA

<213> CORONAVIRUS

```

<400> 1
atattaggtt tttacctacc caggaaaagc caaccaacct cgatctcttg tagatctggt      60
ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgcattgccta gtgcacctac      120
gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct      180
tctgcagact gcttacgggt tcgtccgtgt tgcagtcgat catcagcata cctagggtttc      240
gtccgggtgt gaccgaaaag taagatggag agccttggtc ttggtgtcaa cgagaaaaca      300
cacgtccaac tcagtttgcc tgccttcag gttagagacg tgctagtgcg tggcttcggg      360
gactctgtgg aagaggccct atcggaggca cgtgaacacc tcaaaaatgg cacttgtggt      420
ctagtagagc tggaaaaagg cgtactgccc cagcttgaac agccctatgt gttcattaaa      480
cgttctgatg ccttaagcac caatcacggc cacaaggtcg ttgagctggt tgcagaaatg      540
gacggcattc agtacggtcg tagcgggtata acactgggag tactcgtgcc acatgtgggc      600
gaaaccccaa ttgcataccg caatgttctt cttcgtaaga acggtataaa gggagccggt      660
ggcatagct atggcatcga tctaaagtct tatgacttag gtgacgagct tggcactgat      720
cccattgaag attatgaaca aaactggaac actaagcatg gcagtgggtgc actccgtgaa      780
ctcactcgtg agctcaatgg aggtgcagtc actcgtatg tcgacaacaa tttctgtggc      840
ccagatgggt accctcttga ttgcatcaaa gattttctcg cacgcgcggg caagtcaatg      900
tgcactcttt ccgaacaact tgattacatc gagtcgaaga gaggtgtcta ctgctgccgt      960
gaccatgagc atgaaattgc ctgggttact gagcgtctctg ataagagcta cgagcaccag     1020
acacccttcg aaattaagag tgccaagaaa tttgacactt tcaaagggga atgccccaaag     1080
tttgtgtttc ctcttaactc aaaagtcaaa gtcattcaac cacgtgttga aaagaaaaag     1140
actgaggggt tcatggggcg tatacgtctt gtgtaccctg ttgcatctcc acaggagtgt     1200
aacaatatgc acttgtctac cttgatgaaa tgtaatcatt gcgatgaagt ttcattggcag     1260
acgtgcgact ttctgaaagc cacttgtgaa catttgtgca ctgaaaattt agttattgaa     1320
ggacctacta catgtgggta cctacctact aatgctgtag tgaaaatgcc atgtcctgcc     1380

```



## S226CAS111.ST25

tgtcaagacc	cagagattgg	acctgagcat	agtgttgacg	attatcacia	ccactcaaac	1440
attgaaactc	gactccgcaa	gggaggtagg	actagatggt	ttggaggctg	tgtgtttgcc	1500
tatgttggct	gctataataa	gcgtgcctac	tgggttcctc	gtgctagtgc	tgatattggc	1560
tcaggccata	ctggcattac	tggtgacaat	gtggagacct	tgaatgagga	tctccttgag	1620
atactgagtc	gtgaacgtgt	taacattaac	attgttggcg	atcttcattt	gaatgaagag	1680
gttgccatca	ttttggcatc	tttctctgct	tctacaagtg	cctttattga	cactataaag	1740
agtcttgatt	acaagtcttt	caaaaccatt	gttgagtcct	gcggtaacta	taaagttacc	1800
aagggaaagc	ccgtaaaagg	tgcttggaac	attggacaac	agagatcagt	tttaacacca	1860
ctgtgtggtt	ttccctcaca	ggctgctggt	gttatcagat	caatctttgc	gcgcacactt	1920
gatgcagcaa	accactcaat	tcctgatttg	caaagagcag	ctgtcaccat	acttgatggt	1980
atctctgaac	agtcattacg	tcttgtcgac	gccatgggtt	atacttcaga	cctgctcacc	2040
aacagtgtca	ttattatggc	atatgtaact	ggtggtcctt	tacaacagac	ttctcagtgg	2100
ttgtctaata	ttttgggcac	tactgttgaa	aaactcaggc	ctatctttga	atggattgag	2160
gcgaaactta	gtgcaggagt	tgaatttctc	aaggatgctt	gggagattct	caaatttctc	2220
attacaggtg	tttttgacat	cgtcaagggt	caaatacagg	ttgcttcaga	taacatcaag	2280
gattgtgtaa	aatgcttcat	tgatgttggt	aacaaggcac	tcgaaatgtg	cattgatcaa	2340
gtcactatcg	ctggcgcaaa	gttgcgatca	ctcaacttag	gtgaagtctt	catcgctcaa	2400
agcaagggac	tttaccgtca	gtgtatacgt	ggcaaggagc	agctgcaact	actcatgcct	2460
cttaaggcac	caaaagaagt	aacctttctt	gaagggtgatt	cacatgacac	agtacttacc	2520
tctgaggagg	ttgttctcaa	gaacggtgaa	ctcgaagcac	tcgagacgcc	cgttgatagc	2580
ttcaciaaat	gagctatcgt	tggcacacca	gtctgtgtaa	atggcctcat	gctcttagag	2640
attaaggaca	aagaacaata	ctgcgcattg	tctcctgggt	tactggctac	aaacaatgtc	2700
tttcgcttaa	aaggggggtg	accaattaaa	ggtgtaacct	ttggagaaga	tactgtttgg	2760
gaagttcaag	gttacaagaa	tgtgagaatc	acatttgagc	ttgatgaacg	tgttgacaaa	2820
gtgcttaatg	aaaagtgtct	tgtctacact	gttgaatccg	gtaccgaagt	tactgagttt	2880
gcatgtgttg	tagcagaggc	tgttgtgaag	actttacaac	cagtttctga	tctccttacc	2940
aacatgggta	ttgatcttga	tgagtggagt	gtagctacat	tctacttatt	tgatgatgct	3000
ggtgaagaaa	acttttcata	acgtatgtat	tgttcctttt	accctccaga	tgaggaagaa	3060
gaggacgatg	cagagtgtga	ggaagaagaa	attgatgaaa	cctgtgaaca	tgagtacggt	3120
acagagggat	attatcaagg	tctccctctg	gaatttggtg	cctcagctga	aacagttcga	3180
gttgaggaag	aagaagagga	agactggctg	gatgatacta	ctgagcaatc	agagattgag	3240
ccagaaccag	aacctacacc	tgaagaacca	gttaatcagt	ttactgggtta	tttaaaactt	3300
actgacaatg	ttgccattaa	atgtgttgac	atcgtaagg	aggcaciaag	tgctaatact	3360
atggtgattg	taaatgctgc	taacatacac	ctgaaacatg	gtggtggtgt	agcaggtgca	3420

## S226CAS111.ST25

ctcaacaagg caaccaatgg tgccatgcaa aaggagagtg atgattacat taagctaaat 3480  
 ggccctctta cagtaggagg gtcttggttg ctttctggac ataatcttgc taagaagtgt 3540  
 ctgcatgttg ttggacctaa cctaaatgca ggtgaggaca tccagcttct taaggcagca 3600  
 tatgaaaatt tcaattcaca ggacatctta cttgcacat tggtgtcagc aggcataattt 3660  
 ggtgctaaac cacttcagtc tttaacaagtg tgcgtgcaga cggttcgtac acaggtttat 3720  
 attgcagtca atgacaaagc tctttatgag cagggttgca tggattatct tgataacctg 3780  
 aagcctagag tggaagcacc taaacaagag gagccaccaa acacagaaga ttccaaaact 3840  
 gaggagaaat ctgtcgtaca gaagcctgtc gatgtgaagc caaaaattaa ggcctgcatt 3900  
 gatgaggtta ccacaacact ggaagaaact aagtttctta ccaataagtt actcttgttt 3960  
 gctgatatca atggtaagct ttacatgat tctcagaaca tgcttagagg tgaagatatg 4020  
 tctttccttg agaaggatgc accttacatg gtaggtgatg ttatcactag tggatgatgc 4080  
 acttggtgtg taataccctc caaaaaggct ggtggcacta ctgagatgct ctcaagagct 4140  
 ttgaagaaag tgccagttga tgagtatata accacgtacc ctggacaagg atgtgctggt 4200  
 tatacacttg aggaagctaa gactgctctt aagaaatgca aatctgcatt ttatgtacta 4260  
 ccttcagaag cacctaagtc taaggaagag attctaggaa ctgtatcctg gaatttgaga 4320  
 gaaatgcttg ctcatgctga agagacaaga aaattaatgc ctatatgcat ggatgttaga 4380  
 gccataatgg caaccatcca acgtaagtat aaaggaatta aaattcaaga gggcatcggt 4440  
 gactatggtg tccgattctt cttttatact agtaaagagc ctgtagcttc tattattacg 4500  
 aagctgaact ctctaaatga gccgcttgtc acaatgccaa ttggttatgt gacacatggt 4560  
 tttaatcttg aagaggctgc gcgctgtatg cgttctctta aagctcctgc cgtagtgtca 4620  
 gtatcatcac cagatgctgt tactacatat aatggatacc tcacttcgtc atcaaagaca 4680  
 tctgaggagc actttgtaga aacagtttct ttggctggct cttacagaga ttggtcctat 4740  
 tcaggacagc gtacagagtt aggtgttgaa tttcttaagc gtggtgacaa aattgtgtac 4800  
 cacactctgg agagccccgt cgagtttcat cttgacggtg aggttctttc acttgacaaa 4860  
 ctaaagagtc tcttatccct gcgggaggtt aagactataa aagtgttcac aactgtggac 4920  
 aacactaatc tccacacaca gcttggtgat atgtctatga catatggaca gcagtttggt 4980  
 ccaacatact tggatggtgc tgatgttaca aaaattaaac ctcatgtaaa tcatgagggt 5040  
 aagactttct ttgtactacc tagtgatgac acactacgta gtgaagcttt cgagtactac 5100  
 catactcttg atgagagttt tcttggtagg tacatgtctg ctttaaacca cacaaagaaa 5160  
 tggaaatttc ctcaagttgg tggtttaact tcaattaaat gggctgataa caattgttat 5220  
 ttgtctagtg ttttattagc acttcaacag cttgaagtca aattcaatgc accagcactt 5280  
 caagaggctt attatagagc ccgtgctggt gatgctgcta acttttgtgc actcatactc 5340  
 gcttacagta ataaaactgt tggcgagctt ggtgatgtca gagaaactat gacccatctt 5400  
 ctacagcatg ctaatttgga atctgcaaag cgagttctta atgtggtgtg taaacattgt 5460

## S226CAS111.ST25

ggtcagaaaa ctactacctt aacgggtgta gaagctgtga tgtatatggg tactctatct 5520  
 tatgataatc ttaagacagg tgtttccatt ccatgtgtgt gtggctgtga tgctacacaa 5580  
 tatctagtac aacaagagtc ttcttttggt atgatgtctg caccacctgc tgagtataaa 5640  
 ttacagcaag gtacattctt atgtgcgaat gagtacctg gtaactatca gtgtgggtcat 5700  
 tacactcata taactgctaa ggagaccctc tatcgtattg acggagctca ccttacaaag 5760  
 atgtcagagt acaaaggacc agtgactgat gttttctaca aggaaacatc ttacactaca 5820  
 accatcaagc ctgtgtcgtg taaactcgat ggagttactt acacagagat tgaacaaaaa 5880  
 ttggatgggt attataaaaa ggataatgct tactatacag agcagcctat agaccttgta 5940  
 ccaactcaac cattacaaa tgcgagtttt gataatttca aactcacatg ttctaacaca 6000  
 aaatttgctg atgatttaaa tcaaattgaca ggcttcacaa agccagcttc acgagagcta 6060  
 tctgtcacat tcttcccaga cttgaatggc gatgtagtgg ctattgacta tagacactat 6120  
 tcagcgagtt tcaagaaagg tgctaaatta ctgcataagc caattgtttg gcacattaac 6180  
 caggctacaa ccaagacaac gttcaaacca aacacttggg gtttacgttg tctttggagt 6240  
 acaaagccag tagatacttc aaattcattt gaagttctgg cagtagaaga cacacaagga 6300  
 atggacaatc ttgcttgtga aagtcaacaa cccacctctg aagaagtagt ggaaaatcct 6360  
 accatacaga aggaagtcac agagtgtgac gtgaaaacta ccgaagtgtg aggcaatgtc 6420  
 atacttaaac catcagatga aggtgttaaa gtaacacaag agttaggtca tgaggatctt 6480  
 atggctgctt atgtggaaaa cacaagcatt accattaaga aacctaata gctttcacta 6540  
 gccttagggt taaaaacaat tgccactcat ggtattgctg caattaatag tgttccttgg 6600  
 agtaaaattt tggcttatgt caaacattc ttaggacaag cagcaattac aacatcaaat 6660  
 tgcgctaaga gattagcaca acgtgtgttt aacaattata tgccttatgt gtttacatta 6720  
 ttgttccaat tgtgtacttt tactaaaagt accaattcta gaattagagc ttcactacct 6780  
 acaactattg ctaaaaatag tgtaagagt gttgctaaat tatgtttgga tgccggcatt 6840  
 aattatgtga agtcacccaa attttctaaa ttgttcacaa tcgctatgtg gctattgttg 6900  
 ttaagtattt gcttaggttc tctaactctg gtaactgctg cttttgggtgt actcttatct 6960  
 aattttgggt ctccttctta ttgtaatggc gttagagaat tgtatcttaa ttcgtctaac 7020  
 gttactacta tggatttctg tgaaggttct tttccttgca gcatttggtt aagtggatta 7080  
 gactcccttg attcttatcc agctcttgaa accattcagg tgacgatttc atcgtacaag 7140  
 ctagacttga caattttagg tctggccgct gagtgggttt tggcatatat gttgttcaca 7200  
 aaattctttt atttattagg tctttcagct ataatgcagg tgttctttgg ctattttgct 7260  
 agtcatttca tcagcaattc ttggctcatg tggtttatca ttagtattgt acaaattggca 7320  
 cccgtttctg caatgggttag gatgtacatc ttctttgctt ctttctacta catatggaag 7380  
 agctatgttc atatcatgga tggttgcacc tcttcgactt gcatgatgtg ctataagcgc 7440  
 aatcgtgcca cacgcgttga gtgtacaact attgttaatg gcatgaagag atctttctat 7500

## S226CAS111.ST25

gtctatgcaa atggaggccg tggcttctgc aagactcaca attggaattg tctcaattgt	7560
gacacatttt gactggttag tacattcatt agtgatgaag ttgctcgtga tttgtcactc	7620
cagtttaaaa gaccaatcaa ccctactgac cagtcacgt atattgttga tagtggtgct	7680
gtgaaaaatg gcgcgcttca cctctacttt gacaaggctg gtcaaaagac ctatgagaga	7740
catccgctct cccattttgt caatttagac aatttgagag ctaacaacac taaaggttca	7800
ctgcctatta atgtcatagt ttttgatggc aagtccaaat gcgacgagtc tgcttctaag	7860
tctgcttctg tgtactacag tcagctgatg tgccaaccta ttctgttgct tgaccaagct	7920
cttgatcag acgttggaga tagtactgaa gtttccgtta agatgtttga tgcttatgtc	7980
gacacctttt cagcaacttt tagtgttcct atggaaaaac ttaaggcact tgttgctaca	8040
gctcacagcg agttagcaaa ggggtgtagct ttagatgggtg tcctttctac attcgtgtca	8100
gctgcccagc aagggtgttg tgataccgat gttgacacaa aggatgttat tgaatgtctc	8160
aaactttcac atcactctga cttagaagtg acaggtgaca gttgtaacaa tttcatgctc	8220
acctataata aggttgaaaa catgacgccc agagatcttg gcgcatgtat tgactgtaat	8280
gcaaggcata tcaatgcccc agtagcaaaa agtcacaatg tttcactcat ctggaatgta	8340
aaagactaca tgtctttatc tgaacagctg cgtaaacaaa ttcgtagtgc tgccaagaag	8400
aacaacatac cttttagact aacttgtgct acaactagac aggttgtcaa tgtcataact	8460
actaaaatct cactcaaggg tggttaagatt gttagtactt gttttaaact tatgcttaag	8520
gccacattat tgtgcgttct tgctgcattg gtttgttata tcgttatgcc agtacatata	8580
ttgtcaatcc atgatggtta cacaaatgaa atcattgggtt acaaagccat tcaggatggg	8640
gtcactcgtg acatcatttc tactgatgat tgttttgcaa ataaacatgc tggttttgac	8700
gcatggttta gccagcgtgg tggttcatac aaaaatgaca aaagctgccc tgtagtagct	8760
gctatcatta caagagagat tggtttcata gtgcctggct taccgggtac tgtgctgaga	8820
gcaatcaatg gtgacttctt gcattttcta cctcgtgttt ttagtgctgt tggcaacatt	8880
tgctacacac cttccaaact cattgagtat agtgattttg ctacctctgc ttgcgttctt	8940
gctgctgagt gtacaatttt taaggatgct atgggcaaac ctgtgccata ttgttatgac	9000
actaatttgc tagagggttc tatttcttat agtgagcttc gtccagacac tcgttatgtg	9060
cttatggatg gttccatcat acagtttcct aacacttacc tggagggttc tgttagagta	9120
gtaacaactt ttgatgctga gtactgtaga catggtacat gcgaaaggctc agaagtaggt	9180
atttgccat ctaccagtgg tagatgggtt ctttaataatg agcattacag agctctatca	9240
ggagttttct gtggtgttga tgcgatgaat ctcatagcta acatctttac tcctcttggtg	9300
caacctgtgg gtgctttaga tgtgtctgct tcagtagtgg ctggtggtat tattgccata	9360
ttggtgactt gtgctgccta ctactttatg aaattcagac gtgtttttgg tgagtacaac	9420
catgttggtg ctgctaatact acttttgttt ttgatgtctt tcactatact ctgtctggtg	9480
ccagcttaca gctttctgcc gggagtctac tcagttcttt acttgtactt gacattctat	9540

## S226CAS111.ST25

ttcaccaatg atgtttcatt cttggctcac cttcaatggt ttgccatggt ttctcctatt	9600
gtgccttttt ggataacagc aatctatgta ttctgtatgt ctctgaagca ctgccattgg	9660
ttctttaaca actatcttag gaaaagagtc atgtttaatg gagttacatt tagtaccttc	9720
gaggaggctg ctttgtgtac ctttttgctc aacaaggaaa tgtacctaaa attgcgtagc	9780
gagacactgt tgccacttac acagtataac aggtatcttg ctctatataa caagtacaag	9840
tatttcagtg gagccttaga tactaccagc tatcgtgaag cagcttgctg ccacttagca	9900
aaggctctaa atgactttag caactcaggt gctgatgttc tctaccaacc accacagaca	9960
tcaatcactt ctgctgttct gcagagtggg tttaggaaaa tggcattccc gtcaggcaaa	10020
gttgaagggt gcatggtaca agtaacctgt ggaactacaa ctcttaatgg attgtggttg	10080
gatgacacag tatactgtcc aagacatgtc atttgcacag cagaagacat gcttaatcct	10140
aactatgaag atctgctcat tcgcaaattc aaccatagct ttcttgttca ggctggcaat	10200
gttcaacttc gtgttattgg ccattctatg caaaattgtc tgcttaggct taaagttgat	10260
acttctaacc ctaagacacc caagtataaa tttgtccgta tccaacctgg tcaaacattt	10320
tcagttctag catgctacaa tggttcacca tctgggtgtt atcagtgtgc catgagacct	10380
aatcatacca ttaaagggtc tttccttaat ggatcatgtg gtagtggttg ttttaacatt	10440
gattatgatt gcgtgtcttt ctgctatatg catcatatgg agcttccaac aggagtacac	10500
gctgggtactg acttagaagg taaattctat ggtccatttg ttgacagaca aactgcacag	10560
gctgcaggta cagacacaac cataacatta aatgttttgg catggctgta tgctgctggt	10620
atcaatgggtg ataggtgggt tcttaataga ttcaccacta ctttgaatga ctttaacctt	10680
gtggcaatga agtacaacta tgaacctttg acacaagatc atgttgacat attgggacct	10740
ctttctgctc aaacaggaat tgccgtctta gatatgtgtg ctgctttgaa agagctgctg	10800
cagaatggta tgaatggtcg tactatcctt ggtagcacta ttttagaaga tgagtttaca	10860
ccatttgatg ttgttagaca atgctctggg gttaccttcc aaggtaagtt caagaaaatt	10920
gttaagggca ctcatcattg gatgctttta actttcttga catcactatt gattcttggt	10980
caaagtacac agtggtcact gtttttcttt gtttacgaga atgctttctt gccatttact	11040
cttggtatta tggcaattgc tgcattgtgt atgctgcttg ttaagcataa gcacgcattc	11100
ttgtgcttgt ttctgttacc ttctcttgca acagttgctt actttaatat ggtctacatg	11160
cctgctagct gggatgatgc tatcatgaca tggcttgaat tggctgacac tagcttgtct	11220
ggttataggtc ttaaggattg tgttatgtat gcttcagctt tagttttgct tattctcatg	11280
acagctcgca ctgtttatga tgatgctgct agacgtgttt ggacactgat gaatgtcatt	11340
acacttgttt acaaagtcta ctatggtaat gcttttagatc aagctatttc catgtgggcc	11400
ttagttatgt ctgtaacctc taactattct ggtgtcgtta cgactatcat gtttttagct	11460
agagctatag tgtttggtgt tgttgagtat taccattgtt tatttattac tggcaacacc	11520
ttacagtgtg tcatgcttgt ttattgtttc ttaggctatt gttgctgctg ctactttggc	11580

S226CAS111.ST25

cttttctgtt tactcaaccg ttacttcagg cttactcttg gtgtttatga ctacttggtc 11640  
 tctacacaag aatttaggta tatgaactcc caggggcttt tgcctcctaa gagtagtatt 11700  
 gatgctttca agcttaacat taagttgttg ggtattggag gtaaaccatg tatcaagggt 11760  
 gctactgtac agtctaaaat gtctgacgta aagtgcacat ctgtggtact gctctcgggt 11820  
 cttcaacaac ttagagtaga gtcactcttct aaattgtggg cacaatgtgt acaactccac 11880  
 aatgatattc ttcttgcaaa agacacaact gaagctttcg agaagatggt ttctcttttg 11940  
 tctgttttgc tatccatgca ggggtgctgta gacattaata ggttggtgca ggaaatgctc 12000  
 gataaccgtg ctactcttca ggctattgct tcagaattta gttctttacc atcatatgcc 12060  
 gcttatgcca ctgcccagga ggcctatgag caggctgtag ctaatggtga ttctgaagtc 12120  
 gttctcaaaa agttaagaa atctttgaat gtggctaaat ctgagtttga ccgtgatgct 12180  
 gccatgcaac gcaagttgga aaagatggca gatcaggcta tgacccaaat gtacaaacag 12240  
 gcaagatctg aggacaagag ggcaaaagta actagtgcta tgcaacaat gctcttcact 12300  
 atgcttagga agcttgataa tgatgcactt aacaacatta tcaacaatgc gcgtgatggt 12360  
 tgtgttcac tcaacatcat accattgact acagcagcca aactcatggt tgttgtccct 12420  
 gattatggta cctacaagaa cacttgatggt ggtaacacct ttacatatgc atctgcactc 12480  
 tgggaaatcc agcaagttgt tgatgcggat agcaagattg ttcaacttag tgaaattaac 12540  
 atggacaatt caccaaattt ggcttggcct cttattgtta cagctctaag agccaactca 12600  
 gctgttaaac tacagaataa tgaactgagt ccagtagcac tacgacagat gtcctgtgca 12660  
 gctggtacca cacaaacagc ttgtactgat gacaatgcac ttgcctacta taacaattcg 12720  
 aagggaggta ggtttgtgct ggcattacta tcagaccacc aagatctcaa atgggctaga 12780  
 ttccctaaga gtgatggtac aggtacaatt tacacagaac tggaaccacc ttgtaggttt 12840  
 gttacagaca caccaaaagg gcctaaagt aaatacttgt acttcatcaa aggcttaaac 12900  
 aacctaata gaggtatggt gctgggcagt ttagctgcta cagtacgtct tcaggctgga 12960  
 aatgctacag aagtacctgc caattcaact gtgctttcct tctgtgcttt tgtagtagac 13020  
 cctgctaaag catataagga ttacctagca agtggaggac aaccaatcac caactgtgtg 13080  
 aagatgttgt gtacacacac tggtagagga caggcaatta ctgtaacacc agaagctaac 13140  
 atggaccaag agtccttttg tggtagcttca tgttgtctgt attgtagatg ccacattgac 13200  
 catccaaatc cttaaaggatt ctgtgacttg aaaggtaagt acgtccaaat acctaccact 13260  
 tgtgctaatt acccagtggg ttttactctt agaaacacag tctgtaccgt ctgagggaatg 13320  
 tggaaagggt atggctgtag ttgtgaccaa ctccgcgaac ccttgatgca gtctgcggat 13380  
 gcatcaacgt ttttaaaccg gtttgcggtg taagtgcagc ccgtcttaca ccgtgcggca 13440  
 caggcactag tactgatgtc gtctacaggg cttttgatat ttacaacgaa aaagttgctg 13500  
 gttttgcaaa gttcctaaaa actaattgct gtcgcttcca ggagaaggat gaggaaggca 13560  
 atttattaga ctcttacttt gtagttaaga ggcatactat gtctaactac caacatgaag 13620



S226CAS111.ST25

agactatttta	taacttgggtt	aaagattgtc	cagcggttgc	tgtccatgac	tttttcaagt	13680
ttagagtaga	tggtgacatg	gtaccacata	tatcacgtca	gcgtctaact	aaatacacaa	13740
tggctgattt	agtctatgct	ctacgtcatt	ttgatgaggg	taattgtgat	acattaaaag	13800
aaatactcgt	cacatacaat	tgctgtgatg	atgattatth	caataagaag	gattgggtatg	13860
acttcgtaga	gaatcctgac	atcttacgcg	tatatgctaa	cttaggtgag	cgtgtacgcc	13920
aatcattatt	aaagactgta	caattctgcy	atgctatgcy	tgatgcaggc	attgtaggcy	13980
tactgacatt	agataatcag	gatcttaatg	ggaactggta	cgatttcggt	gatttcgtac	14040
aagtagcacc	aggctgcgga	gttcctattg	tggattcata	ttactcattg	ctgatgcccc	14100
tcctcacttt	gactagggca	ttggctgctg	agtcccatat	ggatgctgat	ctcgcaaaac	14160
cacttattaa	gtgggatttg	ctgaaatatg	atthttacgga	agagagactt	tgtctcttcg	14220
accgttatth	taaatattgg	gaccagacat	accatcccaa	ttgtattaac	tgtttggatg	14280
ataggtgtat	ccttcattgt	gcaaacttht	atgtgttatt	ttctactgtg	ttccaccta	14340
caagtttttg	accactagta	agaaaaatat	ttgtagatgg	tgttccttht	gttgtttcaa	14400
ctggatacca	ttttcgtgag	ttaggagtcg	tacataatca	ggatgtaaac	ttacatagct	14460
cgcgtctcag	tttcaaggaa	cttttagtgt	atgctgctga	tccagctatg	catgcagctt	14520
ctggcaatth	attgctagat	aaacgcacta	catgctthtc	agtagctgca	ctaacaacaa	14580
atgttgctth	tcaaactgtc	aaacccggta	atthttaata	agactthtat	gactttgctg	14640
tgtctaaagg	tttctthtaag	gaaggaagtt	ctgttgaact	aaaacacttc	ttctttgctc	14700
aggatggcaa	cgctgctatc	agtgattatg	actattatcg	ttataatctg	ccaacaatgt	14760
gtgatatcag	acaactccta	ttcgtagtth	aagttgttga	taaatactth	gattgttacg	14820
atggtggctg	tattaatgcc	aaccaagtaa	tcgttaacaa	tctggataaa	tcagctgggt	14880
tcccatttht	taaatggggg	aaggctagac	tttattatga	ctcaatgagt	tatgaggatc	14940
aagatgcact	tttcgcgtat	actaagcgt	atgtcatccc	tactataact	caaatgaatc	15000
ttaagtatgc	cattagtgc	aagaatagag	ctcgcaccgt	agctgggtgc	tctatctgta	15060
gtactatgac	aaatagacag	tttcatcaga	aattattgaa	gtcaatagcc	gccactagag	15120
gagctactgt	ggtaattgga	acaagcaagt	tttacgggtg	ctggcataat	atgttaaaaa	15180
ctgtttacag	tgatgtagaa	actccacacc	ttatgggttg	ggattatcca	aaatgtgaca	15240
gagccatgcc	taacatgctt	aggataatgg	cctctcttgt	tcttgctcgc	aaacataaca	15300
cttgctgtaa	cttatcacac	cgthttctaca	ggtttagctaa	cgagtgtgcy	caagtattaa	15360
gtgagatgg	catgtgtggc	ggctcactat	atgttaaacc	aggtggaaca	tcattccggtg	15420
atgctacaac	tgcttatgct	aatagtgtct	ttaacatttg	tcaagctgtt	acagccaatg	15480
taaatgcact	tctttcaact	gatggtaata	agatagctga	caagtatgtc	cgcaatctac	15540
aacacaggct	ctatgagtgt	ctctatagaa	atagggatgt	tgatcatgaa	ttcgtggatg	15600
agthtttacgc	ttacctgcgt	aaacatttct	ccatgatgat	tctthtctgat	gatgccgttg	15660

## S226CAS111.ST25

tgtgctataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg 15720  
 cagttcttta ttatcaaaat aatgtgttca tgtctgaggc aaaatgttgg actgagactg 15780  
 accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag 15840  
 atgattacgt gtacctgcct taccagatc catcaagaat attaggcgca ggctgttttg 15900  
 tcgatgatat tgtcaaaaca gatggtacac ttatgattga aagggttcgtg tcaactggcta 15960  
 ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcacttgt 16020  
 atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt 16080  
 ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta 16140  
 tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga 16200  
 cttcacttcg ttgcggtgcc tgtattagga gaccattcct atgttgcaag tgctgctatg 16260  
 accatgtcat ttcaacatca cacaattag tgttgtctgt taatccctat gtttgcaatg 16320  
 cccaggttg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt 16380  
 gcaagtcaca taagcctccc attagttttc cattatgtgc taatggtcag gtttttggtt 16440  
 tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgcg atagcaacat 16500  
 gtgattggac taatgctggc gattacatac ttgccaacac ttgtactgag agactcaagc 16560  
 ttttcgcagc agaaacgctc aaagccactg aggaaacatt taagctgtca tatggtattg 16620  
 ccactgtacg cgaagtactc tctgacagag aattgcatct ttcatgggag gttggaaaac 16680  
 ctagaccacc attgaacaga aactatgtct ttactgggta ccgtgtaact aaaaatagta 16740  
 aagtacagat tggagagtac acctttgaaa aagggtgacta tgggtgatgct gttgtgtaca 16800  
 gaggtactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcacactg 16860  
 taatgccact tagtgcacct actctagtgc cacaagagca ctatgtgaga attactggct 16920  
 tgtaccaac actcaacatc tcagatgagt tttctagcaa tgttgcaat tatcaaaagg 16980  
 tcggcatgca aaagtactct acactccaag gaccacctgg tactggtaag agtcattttg 17040  
 ccacggact tgctctctat taccatctg ctcgcatagt gtatacggca tgctctcatg 17100  
 cagctgttga tgccctatgt gaaaaggcat taaaatattt gcccatagat aaatgtagta 17160  
 gaatcatacc tgcgcgtgcg cgcgtagagt gttttgataa attcaaagtg aattcaacac 17220  
 tagaacagta tgttttctgc actgtaaatg cattgccaga aacaactgct gacattgtag 17280  
 tctttgatga aatctctatg gctactaatt atgacttgag tgttgtcaat gctagacttc 17340  
 gtgcaaaaaca ctacgtctat attggcgatc ctgctcaatt accagcccc cgacattgc 17400  
 tgactaaagg cactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa 17460  
 taggtccaga catgttcctt ggaacttgtc gccgttggtc tgctgaaatt gttgacactg 17520  
 tgagtgtttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct 17580  
 tcaaatgtt ctacaaagggt gttattacac atgatgtttc atctgcaatc aacagacctc 17640  
 aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta 17700



## S226CAS111.ST25

tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga 17760  
ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa 17820  
cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaagggca aaaattggca 17880  
ttttgtgcat aatgtctgat agagatcttt atgacaaact gcaatttaca agtctagaaa 17940  
taccacgtcg caatgtggct acattacaag cagaaaatgt aactggactt ttttaaggact 18000  
gtagtaagat cattactggg cttcatccta cacaggcacc tacacacctc agcgttgata 18060  
taaagttcaa gactgaagga ttatgtgttg acataccagg cataccaaag gacatgacct 18120  
accgtagact catctctatg atgggtttca aaatgaatta ccaagtcaat ggttacccta 18180  
atatgtttat caccgcgaa gaagctattc gtcacgttcg tgcgtggatt ggctttgatg 18240  
tagagggctg tcatgcaact agagatgctg tgggtactaa cctacctctc cagctaggat 18300  
tttctacagg tgtaactta gtagctgtac cgactggta tgttgacact gaaaataaca 18360  
cagaattcac cagagttaat gcaaaacctc caccagggtga ccagtttaaa catcttatac 18420  
cactcatgta taaaggcttg ccctggaatg tagtgcgat taagatagta caaatgctca 18480  
gtgatacact gaaaggattg tcagacagag tcgtgttcgt cctttgggcg catggctttg 18540  
agcttacatc aatgaagtac tttgtcaaga ttggacctga aagaacgtgt tgtctgtgtg 18600  
acaaacgtgc aacttgcttt tctacttcat cagatactta tgcctgctgg aatcattctg 18660  
tgggttttga ctatgtctat aaccatttta tgattgatgt tcagcagtgg ggctttacgg 18720  
gtaaccttca gagtaacct gaccaacatt gccagggtaca tggaaatgca catgtggcta 18780  
gttgtgatgc tatcatgact agatgttttag cagtccatga gtgctttgtt aagcgcgttg 18840  
attggctctgt tgaataccct attataggag atgaactgag ggtaattct gcttgcagaa 18900  
aagtacaaca catggttgtg aagtctgcat tgcttgctga taagtttcca gttcttcatg 18960  
acattggaaa tccaaaggct atcaagtgtg tgcctcaggc tgaagtagaa tggaagttct 19020  
acgatgctca gccatgtagt gacaaagctt acaaaataga ggaactcttc tattcttatg 19080  
ctacacatca cgataaattc actgatgggtg tttgtttgtt ttggaattgt aacgttgatc 19140  
gttaccacgc caatgcaatt gtgtgtaggt ttgacacaag agtcttgtca aacttgaact 19200  
taccaggctg tgatgggtgg agtttgatg tgaataagca tgcattccac actccagctt 19260  
tcgataaaag tgcatttact aatttaaagc aattgccttt cttttactat tctgatagtc 19320  
cttgtgagtc tcatggcaaa caagtagtgt cggatattga ttatgttcca ctcaaactg 19380  
ctacgtgtat tacacgatgc aatttaggtg gtgctgtttg cagacaccat gcaaatgagt 19440  
accgacagta cttggatgca tataatatga tgatttctgc tggatttagc ctatggattt 19500  
acaaacaatt tgatacttat aacctgtgga atacatttac caggttacag agtttagaaa 19560  
atgtggctta taatgttgtt aataaaggac actttgatgg acacgccggc gaagcacctg 19620  
tttccatcat taataatgct gtttacacaa aggtagatgg tattgatgtg gagatctttg 19680  
aaaataagac aacacttcct gttaatgttg catttgagct ttgggctaag cgtaacatta 19740

## S226CAS111.ST25

aaccagtgcc agagattaag atactcaata atttgggtgt tgatatcgct gctaatactg 19800  
taatctggga ctacaaaaga gaagccccag cacatgtatc tacaataggt gtctgcacaa 19860  
tgactgacat tgccaagaaa cctactgaga gtgcttgttc ttcacttact gtcttgtttg 19920  
atggtagagt ggaaggacag gtagacctt tttagaaacgc ccgtaatggg gttttaataa 19980  
cagaagggtc agtcaaagg ttaacacctt caaagggacc agcacaagct agcgtcaatg 20040  
gagtcacatt aattggagaa tcagtaaaaa cacagtttaa ctactttaag aaagtagacg 20100  
gcattattca acagttgcct gaaacctact ttactcagag cagagactta gaggatttta 20160  
agcccagatc acaaatggaa actgactttc tcgagctcgc tatggatgaa ttcatacagc 20220  
gatataagct cgagggctat gccttcgaac acatcgttta tggagatttc agtcattggac 20280  
aacttggcgg tcttcattta atgataggct tagccaagcg ctcacaagat tcaccactta 20340  
aattagagga ttttatccct atggacagca cagtgaaaaa ttacttcata acagatgcgc 20400  
aaacagggtc atcaaaatgt gtgtgttctg tgattgatct tttacttgat gactttgtcg 20460  
agataataaa gtcacaagat ttgtcagtga tttcaaaagt ggtcaagggtt acaattgact 20520  
atgctgaaat ttcattcatg ctttggtgta aggatggaca tgttgaaacc ttctacccaa 20580  
aactacaagc aagtcaagcg tggcaaccag gtgttgcgat gcctaacttg tacaagatgc 20640  
aaagaatgct tcttgaaaag tgtgacctt agaattatgg tgaaaatgct gttataccaa 20700  
aaggaataat gatgaatgtc gcaaagtata ctcaactgtg tcaatactta aatacactta 20760  
cttttagctgt accctacaac atgagagtta ttcactttgg tgctggctct gataaaggag 20820  
ttgcaccagg tacagctgtg ctgagacaat ggttgccaac tggcacacta cttgtcgatt 20880  
cagatcttaa tgacttcgtc tccgacgcag attctacttt aattggagac tgtgcaacag 20940  
tacatacggc taataaatgg gaccttatta ttagcgatat gtatgacctt aggaccaaac 21000  
atgtgacaaa agagaatgac tctaagaag ggtttttcac ttatctgtgt ggatttataa 21060  
agcaaaaact agccctgggt ggttctatag ctgtaaagat aacagagcat tcttggaatg 21120  
ctgaccttta caagcttatg ggccatttct catggtggac agcttttggt acaaatgtaa 21180  
atgcatcatc atcggaagca tttttaattg gggctaacta tcttggaag ccgaaggaaac 21240  
aaattgatgg ctataccatg catgctaact acattttctg gaggaacaca aatcctatcc 21300  
agttgtcttc ctattcactc ttgacatga gcaaatttcc tcttaaatta agaggaactg 21360  
ctgtaatgtc tcttaaggag aatcaaatca atgatatgat ttattctctt ctggaaaaag 21420  
gtaggcttat cattagagaa aacaacagag ttgtggtttc aagtgatatt cttgttaaca 21480  
actaaacgaa catgtttatt ttcttattat ttcttactct cactagtggg agtgaccttg 21540  
accggtgcac cacttttgat gatgttcaag ctctaatta cactcaacat acttcatcta 21600  
tgaggggggt ttactatcct gatgaaattt ttagatcaga cactctttat ttaactcagg 21660  
atttatttct tccattttat tctaattgta cagggtttca tactattaat catacgtttg 21720  
gcaaccctgt catacctttt aaggatggta tttattttgc tgccacagag aaatcaaatg 21780

## S226CAS111.ST25

ttgtccgtgg ttgggttttt ggttctacca tgaacaacaa gtcacagtcg gtgattatta 21840  
 ttaacaattc tactaatgtt gttatacgag catgtaactt tgaattgtgt gacaaccctt 21900  
 tctttgctgt ttctaaaccc atgggtacac agacacatac tatgatattc gataatgcat 21960  
 ttaattgcac ttctgagtac atatctgatg ccttttcgct tgatgtttca gaaaagtcag 22020  
 gtaattttaa acacttacga gagtttgtgt ttaaaaataa agatgggttt ctctatgttt 22080  
 ataagggcta tcaacctata gatgtagttc gtgatctacc ttctggtttt aacactttga 22140  
 aacctatttt taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag 22200  
 ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctatttt gttggctatt 22260  
 taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg 22320  
 attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca 22380  
 aaggaattta ccagacctct aatttcaggg ttgttccctc aggagatgtt gtgagattcc 22440  
 ctaatattac aaacttgtgt ccttttggag aggtttttaa tgctactaaa ttcccttctg 22500  
 tctatgcatg ggagagaaaa aaaatttcta attgtgttgc tgattactct gtgctctaca 22560  
 actcaacatt tttttcaacc ttttaagtgc atggcgtttc tgccactaag ttgaatgatc 22620  
 tttgcttctc caatgtctat gcagattctt ttgtagtcaa gggagatgat gtaagacaaa 22680  
 tagcgccagg acaaaactgg gttattgctg attataatta taaattgcca gatgatttca 22740  
 tgggttgtgt ccttgcttgg aatactagga acattgatgc tacttcaact ggtaattata 22800  
 attataaata taggtatctt agacatggca agcttaggcc ctttgagaga gacatatcta 22860  
 atgtgccttt ctcccctgat ggcaaaccct gcaccccacc tgctcttaat tgttattggc 22920  
 cattaatga ttatggtttt tacaccacta ctggcattgg ctaccaacct tacagagttg 22980  
 tagtactttc ttttgaactt ttaaattgcac cggccacggg ttgtggacca aaattatcca 23040  
 ctgaccttat taagaaccag tgtgtcaatt ttaattttaa tggactcact ggtactgggtg 23100  
 tgttactcc ttcttcaaag agatttcaac catttcaaca atttggccgt gatgtttctg 23160  
 atttactga ttccgttcga gatcctaaaa catctgaaat attagacatt tcaccttgct 23220  
 cttttggggg tgtaagtgt attacacctg gaacaaatgc ttcattctgaa gttgctgttc 23280  
 tataatcaaga tgttactgc actgatgttt ctacagcaat tcatgcagat caactcacac 23340  
 cagcttggcg catatattct actggaaaca atgtattcca gactcaagca ggctgtctta 23400  
 taggagctga gcatgtcgac acttcttatg agtgcgacat tcctattgga gctggcattt 23460  
 gtgctagtta ccatacagtt tctttattac gtagtactag ccaaaaatct attgtggctt 23520  
 atactatgtc tttaggtgct gatagttcaa ttgcttactc taataacacc attgctatac 23580  
 ctactaactt ttcaattagc attactacag aagtaatgcc tgtttctatg gctaaaacct 23640  
 ccgtagattg taatatgtac atctgcggag attctactga atgtgctaatt ttgcttctcc 23700  
 aatatggtag cttttgcaca caactaaatc gtgcactctc aggtattgct gctgaacagg 23760  
 atcgcaacac acgtgaagtg ttcgctcaag tcaaacaat gtacaaaacc ccaactttga 23820

## S226CAS111.ST25

aatatatttgg tggttttaat ttttcacaaa tattacctga ccctctaaag ccaactaaga 23880  
ggctcttttat tgaggacttg ctctttaata aggtgacact cgctgatgct ggcttcatga 23940  
agcaatatgg cgaatgccta ggtgatatta atgctagaga tctcatttgt gcgcagaagt 24000  
tcaatggact tacagtgttg ccacctctgc tcaactgatga tatgattgct gcctacactg 24060  
ctgctctagt tagtggtact gccactgctg gatggacatt tggtgctggc gctgctcttc 24120  
aaataccttt tgctatgcaa atggcatata ggttcaatgg cattggagtt acccaaaatg 24180  
ttctctatga gaacaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaattc 24240  
aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gttaaccaga 24300  
atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggt gcaatttcaa 24360  
gtgtgctaaa tgatattctt tcgcgacttg ataaagtcga ggcggaggta caaattgaca 24420  
ggttaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg 24480  
ctgctgaaat cagggcttct gctaattctg ctgctactaa aatgtctgag tgtgttcttg 24540  
gacaatcaaa aagagttgac ttttgtggaa agggctacca ctttatgtcc ttcccacaag 24600  
cagccccgca tgggtgttgc ttctacatg tcacgtatgt gccatcccag gagaggaact 24660  
tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaagggtgtt 24720  
ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa 24780  
ttactacaga caatacattt gtctcaggaa attgtgatgt cgttattggc atcattaaca 24840  
acacagttta tgatcctctg caacctgagc ttgactcatt caaagaagag ctggacaagt 24900  
acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt 24960  
ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggctcgctaaa aatttaaattg 25020  
aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggt 25080  
atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt 25140  
gttgcatgac tagttgttgc agttgcctca aggggtgatg ctcttggtgt tcttgctgca 25200  
agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacacataaa 25260  
cgaacttatg gatttgttta tgagattttt tactcttggg tcaattactg cacagccagt 25320  
aaaaattgac aatgcttctc ctgcaagtac tgttcatgct acagcaacga taccgctaca 25380  
agcctcactc cctttcggat ggcttgttat tggcgttgca tttcttgctg tttttcagag 25440  
cgctaccaa ataattgcgc tcaataaaaag atggcagcta gccctttata agggcttcca 25500  
gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgctgc 25560  
tgcaggatag gaggcgaat tttgtacct ctatgccttg atatattttc tacaatgcat 25620  
caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaaccc 25680  
attactttat gatgccaact actttgtttg ctggcacaca cataactatg actactgtat 25740  
accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc 25800  
aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact caggtgttaa 25860

## S226CAS111.ST25

agactatgtc gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca 25920  
 aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttggtta 25980  
 agacccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc 26040  
 aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga 26100  
 aagtgagtac gaacttatgt actcattcgt ttcggaagaa acaggtagct taatagttaa 26160  
 tagcgtactt ctttttcttg ctttcgtggt attcttgcta gtcacactag ccaccttac 26220  
 tgcgcttcga ttgtgtgctg actgctgcaa tattgttaac gtgagtttag taaaaccaac 26280  
 ggtttacgtc tactcgctg ttaaaaatct gaactcttct gaaggagttc ctgatcttct 26340  
 ggtctaaacg aactaactat tattattatt ctgtttggaa cttaacatt gcttatcatg 26400  
 gcagacaacg gtactattac cgttgaggag cttaacaac tcctggaaca atggaaccta 26460  
 gtaatagggt tcctattcct agcctggatt atgttactac aatttgcta ttctaacgg 26520  
 aacaggtttt tgtacataat aaagcttggt ttcctctggc tcttggtggc agtaacactt 26580  
 gcttgttttg tgcttgctgc tgtctacaga attaattggg tgactggcgg gattgctgatt 26640  
 gcaatggctt gtattgtagg cttagctact tcgttgcttc cttcaggctg 26700  
 tttgctcgta cccgctcaat gtggtcattc aaccagaaa caaacattct tctcaatgtg 26760  
 cctctccggg ggacaattgt gaccagaccg ctcattgaaa gtgaacttgt cattggtgct 26820  
 gtgatcattc gtggtcactt gcgaatggcc ggacactccc tagggcgctg tgacattaag 26880  
 gacctgcaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga 26940  
 gcgtcgcagc gtgtaggcac tgattcaggt tttgctgcat acaaccgcta ccgtattgga 27000  
 aactataaat taaatacaga ccacgccggg agcaacgaca atattgcttt gctagtacag 27060  
 taagtgacaa cagatgtttc atcttggtga cttccagggt acaatagcag agatattgat 27120  
 tatcattatg aggactttca ggattgctat ttggaatctt gacgttataa taagttcaat 27180  
 agtgagacaa ttatttaagc ctctaactaa gaagaattat tcggagttag atgatgaaga 27240  
 acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacattga 27300  
 ttgtatttac atcttgcgag ctatatcact atcaggagtg tgttagaggt acgactgtac 27360  
 tactaaaaga accttgccca tcaggaacat acgagggcaa ttcaccattt caccctcttg 27420  
 ctgacaataa atttgacta acttgacta gcacacactt tgcttttgct tgtgctgacg 27480  
 gtactcgaca tacctatcag ctgcgtgcaa gatcagtttc accaaaactt ttcacagac 27540  
 aagaggaggt tcaacaagag ctctactcgc cactttttct cattgttgct gctctagtat 27600  
 ttttaatact ttgcttcacc attaagagaa agacagaatg aatgagctca ctttaattga 27660  
 cttctatttg tgcttttag ctttctgct attccttggt ttaataatgc ttattatatt 27720  
 ttggttttca ctcgaaatcc aggatctaga agaacttgt accaaagtct aaacgaacat 27780  
 gaaacttctc attgttttga cttgtatttc tctatgcagt tgcataatgca ctgtagtaca 27840  
 gcgctgtgca tctaataaac ctcattgtgt tgaagatcct tgtaaggtag aacactaggg 27900

## S226CAS111.ST25

gtaataactta tagcactgct tggctttgtg ctctaggaaa ggttttacct ttccatagat 27960  
 ggcacactat ggttcaaaca tgcacaccta atgttactat caactgtcaa gatccagctg 28020  
 gtggtgctgct tatagctagg tgttggtacc ttcatgaagg tcaccaaact gctgcattta 28080  
 gagacgtact tgttggttta aataaacgaa caaattaaaa tgtctgataa tggaccccaa 28140  
 tcaaaccaac gtagtgcccc ccgcattaca tttggtggac ccacagattc aactgacaat 28200  
 aaccagaatg gaggacgcaa tggggcaagg ccaaaacagc gccgacccca aggtttaccc 28260  
 aataatactg cgtcttggtt cacagctctc actcagcatg gcaaggagga acttagattc 28320  
 cctcgaggcc agggcgttcc aatcaacacc aatagtggtc cagatgacca aattggctac 28380  
 taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc 28440  
 agatggtact tctattacct aggaactggc ccagaagctt cacttccta cggcgctaac 28500  
 aaagaaggca tcgtatgggt tgcaactgag ggagccttga atacacccaa agaccacatt 28560  
 ggcacccgca atcctaataa caatgctgcc accgtgctac aacttcctca aggaacaaca 28620  
 ttgccaaaag gcttctacgc agagggaaagc agaggcggca gtcaagcctc ttctcgctcc 28680  
 tcatcacgta gtcgcggtaa ttcaagaaat tcaactcctg gcagcagtag gggaaattct 28740  
 cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga 28800  
 ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc 28860  
 actaagaaat ctgctgctga ggcatctaaa aagcctcgcc aaaaacgtac tgccacaaaa 28920  
 cagtacaacg tcaactcaagc atttgggaga cgtggtccag aacaaaccca aggaaatttc 28980  
 ggggaccaag acctaatacag acaaggaact gattacaaac attggccgca aattgcacaa 29040  
 tttgctccaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct 29100  
 tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc 29160  
 aaagacaacg tcatactgct gaacaagcac attgacgcat acaaaacatt cccaccaaca 29220  
 gagcctaaaa aggacaaaaa gaaaaagact gatgaagctc agcctttgcc gcagagacaa 29280  
 aagaagcagc ccactgtgac tcttcttctc gcggctgaca tggatgattt ctccagacaa 29340  
 cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggcataaac actcatgatg 29400  
 accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc 29460  
 tactcttgtg cagaatgaat tctcgtaact aaacagcaca agtaggttta gttaacttta 29520  
 atctcacata gcaatcttta atcaatgtgt aacattaggg aggacttgaa agagccacca 29580  
 cattttcatc gaggccacgc ggagtacgat cgaggggtaca gtgaataatg ctagggagag 29640  
 ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg 29700  
 attttaatag cttcttagga gaatgacaaa aaaaaaaaaa aaaaaa 29746

<210> 2<211> 3945<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (89)..(3853)<223>

<400> 2



S226CAS111.ST25

ttctcttctg gaaaaaggta ggcttatcat tagagaaaac aacagagttg tggtttcaag	60
tgatattctt gttaacaact aaacgaac atg ttt att ttc tta tta ttt ctt	112
Met Phe Ile Phe Leu Leu Phe Leu	
1 5	
act ctc act agt ggt agt gac ctt gac cgg tgc acc act ttt gat gat	160
Thr Leu Thr Ser Gly Ser Asp Leu Asp Arg Cys Thr Thr Phe Asp Asp	
10 15 20	
gtt caa gct cct aat tac act caa cat act tca tct atg agg ggg gtt	208
Val Gln Ala Pro Asn Tyr Thr Gln His Thr Ser Ser Met Arg Gly Val	
25 30 35 40	
tac tat cct gat gaa att ttt aga tca gac act ctt tat tta act cag	256
Tyr Tyr Pro Asp Glu Ile Phe Arg Ser Asp Thr Leu Tyr Leu Thr Gln	
45 50 55	
gat tta ttt ctt cca ttt tat tct aat gtt aca ggg ttt cat act att	304
Asp Leu Phe Leu Pro Phe Tyr Ser Asn Val Thr Gly Phe His Thr Ile	
60 65 70	
aat cat acg ttt ggc aac cct gtc ata cct ttt aag gat ggt att tat	352
Asn His Thr Phe Gly Asn Pro Val Ile Pro Phe Lys Asp Gly Ile Tyr	
75 80 85	
ttt gct gcc aca gag aaa tca aat gtt gtc cgt ggt tgg gtt ttt ggt	400
Phe Ala Ala Thr Glu Lys Ser Asn Val Val Arg Gly Trp Val Phe Gly	
90 95 100	
tct acc atg aac aac aag tca cag tgc gtg att att att aac aat tct	448
Ser Thr Met Asn Asn Lys Ser Gln Ser Val Ile Ile Ile Asn Asn Ser	
105 110 115 120	
act aat gtt gtt ata cga gca tgt aac ttt gaa ttg tgt gac aac cct	496
Thr Asn Val Val Ile Arg Ala Cys Asn Phe Glu Leu Cys Asp Asn Pro	
125 130 135	
ttc ttt gct gtt tct aaa ccc atg ggt aca cag aca cat act atg ata	544
Phe Phe Ala Val Ser Lys Pro Met Gly Thr Gln Thr His Thr Met Ile	
140 145 150	
ttc gat aat gca ttt aat tgc act ttc gag tac ata tct gat gcc ttt	592
Phe Asp Asn Ala Phe Asn Cys Thr Phe Glu Tyr Ile Ser Asp Ala Phe	
155 160 165	
tgc ctt gat gtt tca gaa aag tca ggt aat ttt aaa cac tta cga gag	640
Ser Leu Asp Val Ser Glu Lys Ser Gly Asn Phe Lys His Leu Arg Glu	
170 175 180	
ttt gtg ttt aaa aat aaa gat ggg ttt ctc tat gtt tat aag ggc tat	688
Phe Val Phe Lys Asn Lys Asp Gly Phe Leu Tyr Val Tyr Lys Gly Tyr	
185 190 195 200	
caa cct ata gat gta gtt cgt gat cta cct tct ggt ttt aac act ttg	736
Gln Pro Ile Asp Val Val Arg Asp Leu Pro Ser Gly Phe Asn Thr Leu	
205 210 215	
aaa cct att ttt aag ttg cct ctt ggt att aac att aca aat ttt aga	784
Lys Pro Ile Phe Lys Leu Pro Leu Gly Ile Asn Ile Thr Asn Phe Arg	
220 225 230	
gcc att ctt aca gcc ttt tca cct gct caa gac att tgg ggc acg tca	832
Ala Ile Leu Thr Ala Phe Ser Pro Ala Gln Asp Ile Trp Gly Thr Ser	
235 240 245	
gct gca gcc tat ttt gtt ggc tat tta aag cca act aca ttt atg ctc	880
Ala Ala Ala Tyr Phe Val Gly Tyr Leu Lys Pro Thr Thr Phe Met Leu	

## S226CAS111.ST25

250	255	260	
aag tat gat gaa aat ggt aca atc aca gat gct gtt gat tgt tct caa Lys Tyr Asp Glu Asn Gly Thr Ile Thr Asp Ala Val Asp Cys Ser Gln 265 270 275 280			928
aat cca ctt gct gaa ctc aaa tgc tct gtt aag agc ttt gag att gac Asn Pro Leu Ala Glu Leu Lys Cys Ser Val Lys Ser Phe Glu Ile Asp 285 290 295			976
aaa gga att tac cag acc tct aat ttc agg gtt gtt ccc tca gga gat Lys Gly Ile Tyr Gln Thr Ser Asn Phe Arg Val Val Pro Ser Gly Asp 300 305 310			1024
gtt gtg aga ttc cct aat att aca aac ttg tgt cct ttt gga gag gtt Val Val Arg Phe Pro Asn Ile Thr Asn Leu Cys Pro Phe Gly Glu Val 315 320 325			1072
ttt aat gct act aaa ttc cct tct gtc tat gca tgg gag aga aaa aaa Phe Asn Ala Thr Lys Phe Pro Ser Val Tyr Ala Trp Glu Arg Lys Lys 330 335 340			1120
att tct aat tgt gtt gct gat tac tct gtg ctc tac aac tca aca ttt Ile Ser Asn Cys Val Ala Asp Tyr Ser Val Leu Tyr Asn Ser Thr Phe 345 350 355 360			1168
ttt tca acc ttt aag tgc tat ggc gtt tct gcc act aag ttg aat gat Phe Ser Thr Phe Lys Cys Tyr Gly Val Ser Ala Thr Lys Leu Asn Asp 365 370 375			1216
ctt tgc ttc tcc aat gtc tat gca gat tct ttt gta gtc aag gga gat Leu Cys Phe Ser Asn Val Tyr Ala Asp Ser Phe Val Val Lys Gly Asp 380 385 390			1264
gat gta aga caa ata gcg cca gga caa act ggt gtt att gct gat tat Asp Val Arg Gln Ile Ala Pro Gly Gln Thr Gly Val Ile Ala Asp Tyr 395 400 405			1312
aat tat aaa ttg cca gat gat ttc atg ggt tgt gtc ctt gct tgg aat Asn Tyr Lys Leu Pro Asp Phe Met Gly Cys Val Leu Ala Trp Asn 410 415 420			1360
act agg aac att gat gct act tca act ggt aat tat aat tat aaa tat Thr Arg Asn Ile Asp Ala Thr Ser Thr Gly Asn Tyr Asn Tyr Lys Tyr 425 430 435 440			1408
agg tat ctt aga cat ggc aag ctt agg ccc ttt gag aga gac ata tct Arg Tyr Leu Arg His Gly Lys Leu Arg Pro Phe Glu Arg Asp Ile Ser 445 450 455			1456
aat gtg cct ttc tcc cct gat ggc aaa cct tgc acc cca cct gct ctt Asn Val Pro Phe Ser Pro Asp Gly Lys Pro Cys Thr Pro Pro Ala Leu 460 465 470			1504
aat tgt tat tgg cca tta aat gat tat ggt ttt tac acc act act ggc Asn Cys Tyr Trp Pro Leu Asn Asp Tyr Gly Phe Tyr Thr Thr Thr Gly 475 480 485			1552
att ggc tac caa cct tac aga gtt gta gta ctt tct ttt gaa ctt tta Ile Gly Tyr Gln Pro Tyr Arg Val Val Val Leu Ser Phe Glu Leu Leu 490 495 500			1600
aat gca ccg gcc acg gtt tgt gga cca aaa tta tcc act gac ctt att Asn Ala Pro Ala Thr Val Cys Gly Pro Lys Leu Ser Thr Asp Leu Ile 505 510 515 520			1648
aag aac cag tgt gtc aat ttt aat ttt aat gga ctc act ggt act ggt Lys Asn Gln Cys Val Asn Phe Asn Phe Asn Gly Leu Thr Gly Thr Gly 525 530 535 540			1696



## S226CAS111.ST25

525				530				535								
gtg	tta	act	cct	tct	tca	aag	aga	ttt	caa	cca	ttt	caa	caa	ttt	ggc	1744
Val	Leu	Thr	Pro	Ser	Ser	Lys	Arg	Phe	Gln	Pro	Phe	Gln	Gln	Phe	Gly	
			540					545					550			
cgt	gat	gtt	tct	gat	ttc	act	gat	tcc	gtt	cga	gat	cct	aaa	aca	tct	1792
Arg	Asp	Val	Ser	Asp	Phe	Thr	Asp	Ser	Val	Arg	Asp	Pro	Lys	Thr	Ser	
		555					560					565				
gaa	ata	tta	gac	att	tca	cct	tgc	tct	ttt	ggg	ggt	gta	agt	gta	att	1840
Glu	Ile	Leu	Asp	Ile	Ser	Pro	Cys	Ser	Phe	Gly	Gly	Val	Ser	Val	Ile	
	570					575					580					
aca	cct	gga	aca	aat	gct	tca	tct	gaa	gtt	gct	gtt	cta	tat	caa	gat	1888
Thr	Pro	Gly	Thr	Asn	Ala	Ser	Ser	Glu	Val	Ala	Val	Leu	Tyr	Gln	Asp	
	585				590					595					600	
gtt	aac	tgc	act	gat	gtt	tct	aca	gca	att	cat	gca	gat	caa	ctc	aca	1936
Val	Asn	Cys	Thr	Asp	Val	Ser	Thr	Ala	Ile	His	Ala	Asp	Gln	Leu	Thr	
				605					610					615		
cca	gct	tgg	cgc	ata	tat	tct	act	gga	aac	aat	gta	ttc	cag	act	caa	1984
Pro	Ala	Trp	Arg	Ile	Tyr	Ser	Thr	Gly	Asn	Asn	Val	Phe	Gln	Thr	Gln	
			620					625					630			
gca	ggc	tgt	ctt	ata	gga	gct	gag	cat	gtc	gac	act	tct	tat	gag	tgc	2032
Ala	Gly	Cys	Leu	Ile	Gly	Ala	Glu	His	Val	Asp	Thr	Ser	Tyr	Glu	Cys	
		635					640					645				
gac	att	cct	att	gga	gct	ggc	att	tgt	gct	agt	tac	cat	aca	gtt	tct	2080
Asp	Ile	Pro	Ile	Gly	Ala	Gly	Ile	Cys	Ala	Ser	Tyr	His	Thr	Val	Ser	
	650					655					660					
tta	tta	cgt	agt	act	agc	caa	aaa	tct	att	gtg	gct	tat	act	atg	tct	2128
Leu	Leu	Arg	Ser	Thr	Ser	Gln	Lys	Ser	Ile	Val	Ala	Tyr	Thr	Met	Ser	
	665				670					675				680		
tta	ggt	gct	gat	agt	tca	att	gct	tac	tct	aat	aac	acc	att	gct	ata	2176
Leu	Gly	Ala	Asp	Ser	Ser	Ile	Ala	Tyr	Ser	Asn	Asn	Thr	Ile	Ala	Ile	
				685					690					695		
cct	act	aac	ttt	tca	att	agc	att	act	aca	gaa	gta	atg	cct	gtt	tct	2224
Pro	Thr	Asn	Phe	Ser	Ile	Ser	Ile	Thr	Thr	Glu	Val	Met	Pro	Val	Ser	
			700					705					710			
atg	gct	aaa	acc	tcc	gta	gat	tgt	aat	atg	tac	atc	tgc	gga	gat	tct	2272
Met	Ala	Lys	Thr	Ser	Val	Asp	Cys	Asn	Met	Tyr	Ile	Cys	Gly	Asp	Ser	
		715					720					725				
act	gaa	tgt	gct	aat	ttg	ctt	ctc	caa	tat	ggt	agc	ttt	tgc	aca	caa	2320
Thr	Glu	Cys	Ala	Asn	Leu	Leu	Leu	Gln	Tyr	Gly	Ser	Phe	Cys	Thr	Gln	
	730				735						740					
cta	aat	cgt	gca	ctc	tca	ggt	att	gct	gct	gaa	cag	gat	cgc	aac	aca	2368
Leu	Asn	Arg	Ala	Leu	Ser	Gly	Ile	Ala	Ala	Glu	Gln	Asp	Arg	Asn	Thr	
	745				750					755					760	
cgt	gaa	gtg	ttc	gct	caa	gtc	aaa	caa	atg	tac	aaa	acc	cca	act	ttg	2416
Arg	Glu	Val	Phe	Ala	Gln	Val	Lys	Gln	Met	Tyr	Lys	Thr	Pro	Thr	Leu	
				765					770					775		
aaa	tat	ttt	ggt	ggt	ttt	aat	ttt	tca	caa	ata	tta	cct	gac	cct	cta	2464
Lys	Tyr	Phe	Gly	Gly	Phe	Asn	Phe	Ser	Gln	Ile	Leu	Pro	Asp	Pro	Leu	
			780					785					790			
aag	cca	act	aag	agg	tct	ttt	att	gag	gac	ttg	ctc	ttt	aat	aag	gtg	2512
Lys	Pro	Thr	Lys	Arg	Ser	Phe	Ile	Glu	Asp	Leu	Leu	Phe	Asn	Lys	Val	

## S226CAS111.ST25

795	800	805	
aca ctc gct gat gct ggc ttc atg aag caa tat ggc gaa tgc cta ggt Thr Leu Ala Asp Ala Gly Phe Met Lys Gln Tyr Gly Glu Cys Leu Gly 810 815 820			2560
gat att aat gct aga gat ctc att tgt gcg cag aag ttc aat gga ctt Asp Ile Asn Ala Arg Asp Leu Ile Cys Ala Gln Lys Phe Asn Gly Leu 825 830 835			2608
aca gtg ttg cca cct ctg ctc act gat gat atg att gct gcc tac act Thr Val Leu Pro Pro Leu Leu Thr Asp Asp Met Ile Ala Ala Tyr Thr 845 850 855			2656
gct gct cta gtt agt ggt act gcc act gct gga tgg aca ttt ggt gct Ala Ala Leu Val Ser Gly Thr Ala Thr Ala Gly Trp Thr Phe Gly Ala 860 865 870			2704
ggc gct gct ctt caa ata cct ttt gct atg caa atg gca tat agg ttc Gly Ala Ala Leu Gln Ile Pro Phe Ala Met Gln Met Ala Tyr Arg Phe 875 880 885			2752
aat ggc att gga gtt acc caa aat gtt ctc tat gag aac caa aaa caa Asn Gly Ile Gly Val Thr Thr Asn Val Leu Tyr Glu Asn Gln Lys Gln 890 895 900			2800
atc gcc aac caa ttt aac aag gcg att agt caa att caa gaa tca ctt Ile Ala Asn Gln Phe Asn Lys Ala Ile Ser Gln Ile Gln Glu Ser Leu 905 910 915 920			2848
aca aca aca tca act gca ttg ggc aag ctg caa gac gtt gtt aac cag Thr Thr Thr Ser Thr Ala Leu Gly Lys Leu Gln Asp Val Val Asn Gln 925 930 935			2896
aat gct caa gca tta aac aca ctt gtt aaa caa ctt agc tct aat ttt Asn Ala Gln Ala Leu Asn Thr Leu Val Lys Gln Leu Ser Ser Asn Phe 940 945 950			2944
ggt gca att tca agt gtg cta aat gat atc ctt tcg cga ctt gat aaa Gly Ala Ile Ser Ser Val Leu Asn Asp Ile Leu Ser Arg Leu Asp Lys 955 960 965			2992
gtc gag gcg gag gta caa att gac agg tta att aca ggc aga ctt caa Val Glu Ala Glu Val Gln Ile Asp Arg Leu Ile Thr Gly Arg Leu Gln 970 975 980			3040
agc ctt caa acc tat gta aca caa caa cta atc agg gct gct gaa atc Ser Leu Gln Thr Tyr Val Thr Gln Gln Leu Ile Arg Ala Ala Glu Ile 985 990 995 1000			3088
agg gct tct gct aat ctt gct gct act aaa atg tct gag tgt gtt Arg Ala Ser Ala Asn Leu Ala Ala Thr Lys Met Ser Glu Cys Val 1005 1010 1015			3133
ctt gga caa tca aaa aga gtt gac ttt tgt gga aag ggc tac cac Leu Gly Gln Ser Lys Arg Val Asp Phe Cys Gly Lys Gly Tyr His 1020 1025 1030			3178
ctt atg tcc ttc cca caa gca gcc ccg cat ggt gtt gtc ttc cta Leu Met Ser Phe Pro Gln Ala Ala Pro His Gly Val Val Phe Leu 1035 1040 1045			3223
cat gtc acg tat gtg cca tcc cag gag agg aac ttc acc aca gcg His Val Thr Tyr Val Pro Ser Gln Glu Arg Asn Phe Thr Thr Ala 1050 1055 1060			3268
cca gca att tgt cat gaa ggc aaa gca tac ttc cct cgt gaa ggt Pro Ala Ile Cys His Glu Gly Lys Ala Tyr Phe Pro Arg Glu Gly 1060			3313

## S226CAS111.ST25

1065					1070					1075					
gtt	ttt	gtg	ttt	aat	ggc	act	tct	tgg	ttt	att	aca	cag	agg	aac	3358
Val	Phe	Val	Phe	Asn	Gly	Thr	Ser	Trp	Phe	Ile	Thr	Gln	Arg	Asn	
				1080					1085					1090	
ttc	ttt	tct	cca	caa	ata	att	act	aca	gac	aat	aca	ttt	gtc	tca	3403
Phe	Phe	Ser	Pro	Gln	Ile	Ile	Thr	Thr	Asp	Asn	Thr	Phe	Val	Ser	
				1095					1100					1105	
gga	aat	tgt	gat	gtc	gtt	att	ggc	atc	att	aac	aac	aca	gtt	tat	3448
Gly	Asn	Cys	Asp	Val	Val	Ile	Gly	Ile	Ile	Asn	Asn	Thr	Val	Tyr	
				1110					1115					1120	
gat	cct	ctg	caa	cct	gag	ctt	gac	tca	ttc	aaa	gaa	gag	ctg	gac	3493
Asp	Pro	Leu	Gln	Pro	Glu	Leu	Asp	Ser	Phe	Lys	Glu	Glu	Leu	Asp	
				1125					1130					1135	
aag	tac	ttc	aaa	aat	cat	aca	tca	cca	gat	gtt	gat	ctt	ggc	gac	3538
Lys	Tyr	Phe	Lys	Asn	His	Thr	Ser	Pro	Asp	Val	Asp	Leu	Gly	Asp	
				1140					1145					1150	
att	tca	ggc	att	aac	gct	tct	gtc	gtc	aac	att	caa	aaa	gaa	att	3583
Ile	Ser	Gly	Ile	Asn	Ala	Ser	Val	Val	Asn	Ile	Gln	Lys	Glu	Ile	
				1155					1160					1165	
gac	cgc	ctc	aat	gag	gtc	gct	aaa	aat	tta	aat	gaa	tca	ctc	att	3628
Asp	Arg	Leu	Asn	Glu	Val	Ala	Lys	Asn	Leu	Asn	Glu	Ser	Leu	Ile	
				1170					1175					1180	
gac	ctt	caa	gaa	ttg	gga	aaa	tat	gag	caa	tat	att	aaa	tgg	cct	3673
Asp	Leu	Gln	Glu	Leu	Gly	Lys	Tyr	Glu	Gln	Tyr	Ile	Lys	Trp	Pro	
				1185					1190					1195	
tgg	tat	gtt	tgg	ctc	ggc	ttc	att	gct	gga	cta	att	gcc	atc	gtc	3718
Trp	Tyr	Val	Trp	Leu	Gly	Phe	Ile	Ala	Gly	Leu	Ile	Ala	Ile	Val	
				1200					1205					1210	
atg	gtt	aca	atc	ttg	ctt	tgt	tgc	atg	act	agt	tgt	tgc	agt	tgc	3763
Met	Val	Thr	Ile	Leu	Leu	Cys	Cys	Met	Thr	Ser	Cys	Cys	Ser	Cys	
				1215					1220					1225	
ctc	aag	ggt	gca	tgc	tct	tgt	ggt	tct	tgc	tgc	aag	ttt	gat	gag	3808
Leu	Lys	Gly	Ala	Cys	Ser	Cys	Gly	Ser	Cys	Cys	Lys	Phe	Asp	Glu	
				1230					1235					1240	
gat	gac	tct	gag	cca	gtt	ctc	aag	ggt	gtc	aaa	tta	cat	tac	aca	3853
Asp	Asp	Ser	Glu	Pro	Val	Leu	Lys	Gly	Val	Lys	Leu	His	Tyr	Thr	
				1245					1250					1255	
taaacgaact tatggatttg tttatgagat tttttactct tggatcaatt actgcacagc															3913
cagtaaaaat tgacaatgct tctcctgcaa gt															3945
<210> 3<211> 1255<212> PRT<213> CORONAVIRUS<400> 3															
Met	Phe	Ile	Phe	Leu	Leu	Phe	Leu	Thr	Leu	Thr	Ser	Gly	Ser	Asp	Leu
1				5					10					15	
Asp	Arg	Cys	Thr	Thr	Phe	Asp	Asp	Val	Gln	Ala	Pro	Asn	Tyr	Thr	Gln
			20					25					30		
His	Thr	Ser	Ser	Met	Arg	Gly	Val	Tyr	Tyr	Pro	Asp	Glu	Ile	Phe	Arg
		35					40					45			

## S226CAS111.ST25

Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser  
 50 55 60  
 Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val  
 65 70 75 80  
 Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn  
 85 90 95  
 Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln  
 100 105 110  
 Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys  
 115 120 125  
 Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met  
 130 135 140  
 Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr  
 145 150 155 160  
 Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser  
 165 170 175  
 Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly  
 180 185 190  
 Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp  
 195 200 205  
 Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu  
 210 215 220  
 Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro  
 225 230 235 240  
 Ala Gln Asp Ile Trp Gly Thr Ser Ala Ala Ala Tyr Phe Val Gly Tyr  
 245 250 255  
 Leu Lys Pro Thr Thr Phe Met Leu Lys Tyr Asp Glu Asn Gly Thr Ile  
 260 265 270  
 Thr Asp Ala Val Asp Cys Ser Gln Asn Pro Leu Ala Glu Leu Lys Cys  
 275 280 285  
 Ser Val Lys Ser Phe Glu Ile Asp Lys Gly Ile Tyr Gln Thr Ser Asn  
 290 295 300  
 Phe Arg Val Val Pro Ser Gly Asp Val Val Arg Phe Pro Asn Ile Thr  
 305 310 315 320

## S226CAS111.ST25

Asn Leu Cys Pro Phe Gly Glu Val Phe Asn Ala Thr Lys Phe Pro Ser  
325 330 335

Val Tyr Ala Trp Glu Arg Lys Lys Ile Ser Asn Cys Val Ala Asp Tyr  
340 345 350

Ser Val Leu Tyr Asn Ser Thr Phe Phe Ser Thr Phe Lys Cys Tyr Gly  
355 360 365

Val Ser Ala Thr Lys Leu Asn Asp Leu Cys Phe Ser Asn Val Tyr Ala  
370 375 380

Asp Ser Phe Val Val Lys Gly Asp Asp Val Arg Gln Ile Ala Pro Gly  
385 390 395 400

Gln Thr Gly Val Ile Ala Asp Tyr Asn Tyr Lys Leu Pro Asp Asp Phe  
405 410 415

Met Gly Cys Val Leu Ala Trp Asn Thr Arg Asn Ile Asp Ala Thr Ser  
420 425 430

Thr Gly Asn Tyr Asn Tyr Lys Tyr Arg Tyr Leu Arg His Gly Lys Leu  
435 440 445

Arg Pro Phe Glu Arg Asp Ile Ser Asn Val Pro Phe Ser Pro Asp Gly  
450 455 460

Lys Pro Cys Thr Pro Pro Ala Leu Asn Cys Tyr Trp Pro Leu Asn Asp  
465 470 475 480

Tyr Gly Phe Tyr Thr Thr Thr Gly Ile Gly Tyr Gln Pro Tyr Arg Val  
485 490 495

Val Val Leu Ser Phe Glu Leu Leu Asn Ala Pro Ala Thr Val Cys Gly  
500 505 510

Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn  
515 520 525

Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg  
530 535 540

Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp  
545 550 555 560

Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys  
565 570 575

Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser  
580 585 590

## S226CAS111.ST25

Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr  
 595 600 605  
 Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr  
 610 615 620  
 Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu  
 625 630 635 640  
 His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile  
 645 650 655  
 Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys  
 660 665 670  
 Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala  
 675 680 685  
 Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile  
 690 695 700  
 Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys  
 705 710 715 720  
 Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu  
 725 730 735  
 Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile  
 740 745 750  
 Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys  
 755 760 765  
 Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe  
 770 775 780  
 Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile  
 785 790 795 800  
 Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met  
 805 810 815  
 Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile  
 820 825 830  
 Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr  
 835 840 845  
 Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala  
 850 855 860

S226CAS111.ST25

Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe  
 865 870 875 880  
 Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn  
 885 890 895  
 Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala  
 900 905 910  
 Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly  
 915 920 925  
 Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu  
 930 935 940  
 Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn  
 945 950 955 960  
 Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp  
 965 970 975  
 Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln  
 980 985 990  
 Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala  
 995 1000 1005  
 Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp  
 1010 1015 1020  
 Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala  
 1025 1030 1035  
 Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln  
 1040 1045 1050  
 Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys  
 1055 1060 1065  
 Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser  
 1070 1075 1080  
 Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr  
 1085 1090 1095  
 Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly  
 1100 1105 1110  
 Ile Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp  
 1115 1120 1125

## S226CAS111.ST25

Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser  
1130 1135 1140

Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val  
1145 1150 1155

Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys  
1160 1165 1170

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr  
1175 1180 1185

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe Ile  
1190 1195 1200

Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Leu Leu Cys Cys  
1205 1210 1215

Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Ala Cys Ser Cys Gly  
1220 1225 1230

Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys  
1235 1240 1245

Gly Val Lys Leu His Tyr Thr  
1250 1255

```
<210> 4<211> 3943<212> DNA<213> CORONAVIRUS<400> 4
ctcttctgga aaaaggtagg cttatcatta gagaaaacaa cagagttgtg gtttcaagtg 60
atattcttgt taacaactaa acgaacatgt ttattttctt attatttctt actctcacta 120
gtggtagtga ccttgaccgg tgcaccactt ttgatgatgt tcaagctcct aattacactc 180
aacatacttc atctatgagg ggggtttact atcctgatga aatttttaga tcagacactc 240
tttatttaac tcaggattta tttcttccat tttattctaa tggtacaggg tttcatacta 300
ttaatcatac gtttggcaac cctgtcatac cttttaagga tgggtatttat tttgctgcca 360
cagagaaatc aaatgttgtc cgtggttggg tttttggttc taccatgaac aacaagtcac 420
agtcggtgat tattattaac aattctacta atgttggtat acgagcatgt aactttgaat 480
tgtgtgacaa ccctttcttt gctgtttcta aacccatggg tacacagaca catactatga 540
tattcgataa tgcatttaat tgcactttcg agtacatatc tgatgccttt tcgcttgatg 600
tttcagaaaa gtcaggtaat tttaaacact tacgagagtt tgtgtttaaa aataaagatg 660
ggtttctcta tgtttataag ggctatcaac ctatagatgt agttcgtgat ctaccttctg 720
gttttaacac tttgaaacct atttttaagt tgcctcttgg tattaacatt acaaatttta 780
gagccattct tacagccttt tcacctgctc aagacatttg gggcacgtca gctgcagcct 840
attttgttgg ctatttaaag ccaactacat ttatgctcaa gtatgatgaa aatggtacaa 900
```



S226CAS111.ST25

tcacagatgc tgttgattgt tctcaaaatc cacttgctga actcaaatgc tctgttaaga	960
gctttgagat tgacaaagga atttaccaga cctctaattt cagggttggt ccctcaggag	1020
atgttgtag attccctaatt attacaaact tgtgtccttt tggagagggt tttaatgcta	1080
ctaaattccc ttctgtctat gcatgggaga gaaaaaaaaat ttctaattgt gttgctgatt	1140
actctgtgct ctacaactca acatTTTTTT caacctttaa gtgctatggc gtttctgcc	1200
ctaagttgaa tgatctttgc ttctccaatg tctatgcaga ttcttttgta gtcaaggag	1260
atgatgtaag acaaatacg ccaggacaaa ctggtgttat tgctgattat aattataaat	1320
tgccagatga tttcatgggt tgtgtccttg cttggaatac taggaacatt gatgctactt	1380
caactggtaa ttataattat aaatatagggt atcttagaca tggcaagctt aggcccttg	1440
agagagacat atctaattgt ctttctccc ctgatggcaa accttgacc ccacctgctc	1500
ttaattgtta ttggcatta aatgattatg gtttttacac cactactggc attggctacc	1560
aaccttacag agttgtagta ctttcttttg aacttttaaa tgcaccggcc acggtttgtg	1620
gaccaaatt atccactgac cttattaaga accagtgtgt caattttaat tttaatggac	1680
tactggtac tgggtgtgta actccttctt caaagagatt tcaaccattt caacaatttg	1740
gccgtgatgt ctctgatttc actgattccg ttcgagatcc taaaacatct gaaatattag	1800
acatttcacc ttgctctttt gggggtgtaa gtgtaattac acctggaaca aatgcttcat	1860
ctgaagttgc tgttctatat caagatgtta actgcactga tgtttctaca gcaatccatg	1920
cagatcaact cacaccagct tggcgcatat attctactgg aaacaatgta ttccagactc	1980
aagcaggctg tcttatagga gctgagcatg tgcacacttc ttatgagtgc gacattccta	2040
ttggagctgg catttgctgct agttaccata cagtttcttt attacgtagt actagccaaa	2100
aatctattgt ggcttatact atgtcttttag gtgctgatag ttcaattgct tactctaata	2160
acaccattgc tatacctact aacttttcaa ttagcattac tacagaagta atgcctgttt	2220
ctatggctaa aacctccgta gattgtaata tgtacatctg cggagattct actgaatgtg	2280
ctaatttgct tctccaatat ggtagctttt gcacacaact aaatcgtgca ctctcaggta	2340
ttgctgctga acaggatcgc aacacacgtg aagtgttcgc tcaagtcaaa caaatgtaca	2400
aaacccaac tttgaaatat tttggtggtt ttaatttttc acaaatatta cctgaccctc	2460
taaagccaac taagaggctt tttattgagg acttgctctt taataagggtg acactcgtg	2520
atgctggctt catgaagcaa tatggcgaat gcctagggtga tattaatgct agagatctca	2580
tttgctgcga gaagttcaat gggcttacag tgttgccacc tctgctcact gatgatatga	2640
ttgctgccta cactgctgct ctagttagtg gtactgccac tgctggatgg acatttggtg	2700
ctggcgctgc tcttcaaata ctttttgcta tgcaaatggc atatagggtt aatggcattg	2760
gagttacca aatgtttctc tatgagaacc aaaaacaaat cgccaaccaa tttacaagg	2820
cgattagtca aattcaagaa tcacttaca caacatcaac tgcattgggc aagctgcaag	2880
acgttggtta ccagaatgct caagcattaa acacacttgt taaacaactt agctctaatt	2940

## S226CAS111.ST25

ttggtgcaat ttcaagtgtg ctaaattgata tccttttcgcg acttgataaa gtcgaggcgg 3000  
 aggtacaaat tgacaggcta attacaggca gacttcaaag ccttcaaacc tatgtaacac 3060  
 aacaactaat cagggctgct gaaatcaggg cttctgctaa tcttgctgct actaaaatgt 3120  
 ctgagtgtgt tcttggacaa tcaaaaagag ttgacttttg tggaaagggc taccacctta 3180  
 tgtccttccc acaagcagcc ccgcatgggtg ttgtcttcct acatgtcacg tatgtgccat 3240  
 cccaggagag gaacttcacc acagcgccag caatttgtca tgaaggcaaa gcatacttcc 3300  
 ctctggaagg tgtttttgtg tttaatggca cttcttgggtt tattacacag aggaacttct 3360  
 tttctccaca aataattact acagacaata catttgtctc aggaaattgt gatgtcggtta 3420  
 ttggcatcat taacaacaca gtttatgatc ctctgcaacc tgagcttgac tcattcaaag 3480  
 aagagctgga caagtacttc aaaaatcata catcaccaga tgttgatctt ggcgacattt 3540  
 caggcattaa cgcttctgtc gtcaacattc aaaaagaaat tgaccgcctc aatgagggtcg 3600  
 ctaaaaattt aatatgaatca ctcatgacc ttcaagaatt gggaaaatat gagcaatata 3660  
 ttaaatggcc ttggtatgtt tggctcggct tcattgctgg actaattgcc atcgatcatgg 3720  
 ttacaatctt gctttgttgc atgactagtt gttgcagttg cctcaagggt gcatgctctt 3780  
 gtggttcttg ctgcaagttt gatgaggatg actctgagcc agttctcaag ggtgtcaaat 3840  
 tacattacac ataaacgaac ttatggattt gtttatgaga ttttttactc ttggatcaat 3900  
 tactgcacag ccagtaaaaa ttgacaatgc ttctcctgca agt 3943

<210> 5<211> 2049<212> DNA<213> CORONAVIRUS<400> 5  
 ctcttctgga aaaaggtagg cttatcatta gagaaaacaa cagagttgtg gtttcaagtg 60  
 atattcttgt taacaactaa acgaacatgt ttattttctt attatttctt actctcacta 120  
 gtggtagtga ccttgaccgg tgcaccactt ttgatgatgt tcaagctcct aattacactc 180  
 aacatacttc atctatgagg ggggtttact atcctgatga aatttttaga tcagacactc 240  
 tttatttaac tcaggattta tttcttccat tttattctaa tgttacaggg tttcatacta 300  
 ttaatcatac gtttggcaac cctgtcatac cttttaagga tggatattat tttgctgcca 360  
 cagagaaatc aaatgttgtc cgtggttggg ttttgggttc taccatgaac aacaagtcac 420  
 agtcggtgat tattattaac aattctacta atgttggtat acgagcatgt aactttgaat 480  
 tgtgtgacaa ccctttcttt gctgtttcta aacctatggg tacacagaca catactatga 540  
 tattcgataa tgcatttaat tgcactttcg agtacatata tgatgccttt tcgcttgatg 600  
 tttcagaaaa gtcaggtaat tttaaacact tacgagagtt tgtgtttaaa aataaagatg 660  
 ggtttctcta tgtttataag ggctatcaac ctatagatgt agttcgtgat ctaccttctg 720  
 gttttaacac tttgaaacct atttttaagt tgcctcttgg tattaacatt acaaatttta 780  
 gagccattct tacagccttt tcacctgtc aagacatttg gggcacgtca gctgcagcct 840  
 attttgttgg ctatttaaag ccaactacat ttatgctcaa gtatgatgaa aatggtacaa 900  
 tcacagatgc tgttgattgt tctcaaaatc cacttgcgtga actcaaatgc tctgttaaga 960

S226CAS111.ST25

gctttgagat tgacaaagga atttaccaga cctctaattt cagggttggt ccctcaggag	1020
atgttgtgag attccctaatt attacaaact tgtgtccttt tggagagggt tttaattgcta	1080
ctaaattccc ttctgtctat gcatgggaga gaaaaaaaaat ttctaattgt gttgctgatt	1140
actctgtgct ctacaactca acatTTTTTTT caacctttta gtgctatggc gtttctgcca	1200
ctaagttgaa tgatctttgc ttctccaatg tctatgcaga ttcttttgta gtcaagggag	1260
atgatgtaag acaaatagcg ccaggacaaa ctggtgttat tgctgattat aattataaat	1320
tgccagatga ttcatgggt tgtgtccttg cttggaatac taggaacatt gatgctactt	1380
caactggtaa ttataattat aaatataggt atcttagaca tggcaagctt aggccctttg	1440
agagagacat atctaattgt ctttctccc ctgatggcaa accttgcacc ccacctgctc	1500
ttaattgtta ttggcatta aatgattatg gtttttacac cactactggc attggctacc	1560
aaccttacag agttgtagta ctttcttttg aacttttaaa tgcaccggcc acggtttgtg	1620
gacaaaatt atccactgac cttattaaga accagtgtgt caattttaat tttaatggac	1680
tcactggtac tgggtgtgta actccttctt caaagagatt tcaaccattt caacaatttg	1740
gccgtgatgt ctctgatttc actgattccg ttcgagatcc taaaacatct gaaatattag	1800
acatttcacc ttgctctttt gggggtgtaa gtgtaattac acctggaaca aatgcttcat	1860
ctgaagttgc tgttctatat caagatgtta actgcactga tgtttctaca gcaatccatg	1920
cagatcaact cacaccagct tggcgcatat attctactgg aaacaatgta ttccagactc	1980
aagcaggctg tcttatagga gctgagcatg tcgacacttc ttatgagtgc gacattccta	2040
ttggagctg	2049

<210> 6<211> 2027<212> DNA<213> CORONAVIRUS<400> 6	
catgcagatc aactcacacc agcttggcgc atatattcta ctggaaacaa tgtattccag	60
actcaagcag gctgtcttat aggagctgag catgtcgaca cttcttatga gtgcgacatt	120
cctattggag ctggcatttg tgctagttac catacagttt ctttattacg tagtactagc	180
caaaaatcta ttgtggctta tactatgtct ttaggtgctg atagttcaat tgcttactct	240
aataacacca ttgctatacc tactaacttt tcaattagca ttactacaga agtaatgcct	300
gtttctatgg ctaaaacctc cgtagattgt aatatgtaca tctgcggaga ttctactgaa	360
tgtgctaatt tgcttctcca atatggtagc ttttgcacac aactaaatcg tgcactctca	420
ggtattgctg ctgaacagga tcgcaacaca cgtgaagtgt tcgctcaagt caaacaatg	480
tacaaaaccc caactttgaa atattttggt ggttttaatt ttccacaaat attacctgac	540
cctctaaagc caactaagag gtcttttatt gaggacttgc tctttaataa ggtgacactc	600
gctgatgctg gcttcatgaa gcaatatggc gaatgcctag gtgatattaa tgctagagat	660
ctcattttgtg cgcagaagtt caatgggctt acagtgttgc cacctctgct cactgatgat	720
atgattgctg cctacactgc tgctctagtt agtggtagct ccactgctgg atggacattt	780
ggtgctggcg ctgctcttca aatacctttt gctatgcaaa tggcatatag gttcaatggc	840

S226CAS111.ST25

attggagtta cccaaaatgt tctctatgag aaccaaaaac aaatcgccaa ccaatttaac	900
aaggcgatta gtcaaattca agaatcactt acaacaacat caactgcatt gggcaagctg	960
caagacgttg ttaaccagaa tgctcaagca ttaaacacac ttgttaaaca acttagctct	1020
aattttggtg caatttcaag tgtgctaaat gatatccttt cgcgacttga taaagtcgag	1080
gcgagggtac aaattgacag gttaattaca ggcagacttc aaagccttca aacctatgta	1140
acacaacaac taatcagggc tgctgaaatc agggcttctg ctaatcttgc tgctactaaa	1200
atgtctgagt gtgttcttgg acaatcaaaa agagttgact tttgtggaaa gggctaccac	1260
cttatgtcct tcccacaagc agccccgcac ggtgttgcct tcctacatgt cacgtatgtg	1320
ccatcccagg agaggaactt caccacagcg ccagcaattt gtcatgaagg caaagcatac	1380
ttccctcgtg aaggtgtttt tgtgtttaat ggcacttctt ggtttattac acagaggaac	1440
ttcttttctc cacaaataat tactacagac aatacatttg tctcaggaaa ttgtgatgtc	1500
gttattggcg tcattaacaa cacagtttat gatcctctgc aacctgagct tgactcattc	1560
aaagaagagc tggacaagta cttcaaaaat catacatcac cagatgttga tcttggcgac	1620
atctcaggca ttaacgcttc tgcgtcaac attcaaaaag aaattgaccg cctcaatgag	1680
gtcgctaaaa atttaaatga atcactcatt gaccttcaag aattgggaaa atatgagcaa	1740
tatattaaat ggccttggtg tgtttggctc ggcttcattg ctggactaat tgccatcgtc	1800
atggttacaa tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc	1860
tcttgtgggt cttgctgcaa gtttgatgag gatgactctg agccagttct caaggggtgc	1920
aaattacatt acacataaac gaacttatgg atttgtttat gagatttttt actcttggat	1980
caattactgc acagccagta aaaattgaca atgcttctcc tgcaagt	2027

<210> 7<211> 1096<212> DNA<213> CORONAVIRUS<400> 7	
tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc tcttgtgggt	60
cttgtctgcaa gtttgatgag gatgactctg agccagttct caaggggtgc aaattacatt	120
acacataaac gaacttatgg atttgtttat gagatttttt actcttggat caattactgc	180
acagccagta aaaattgaca atgcttctcc tgcaagtact gttcatgcta cagcaacgat	240
accgctacaa gcctcactcc ctttcggatg gcttgttatt ggcgttgcat ttcttgcgtg	300
ttttcagagc gctaccaaaa taattgcgct caataaaaga tggcagctag ccctttataa	360
gggcttccag ttcatattgca atttactgct gctatttgtt accatctatt cacatctttt	420
gcttgtcgct gcaggtatgg aggcgcaatt tttgtacctc taigccttga tatattttct	480
acaatgcac aacgcatgta gaattattat gagatgttgg ctttgttgga agtgcaaatc	540
caagaacca ttactttatg atgccaacta ctttgtttgc tggcacacac ataactatga	600
ctactgtata ccatataaca gtgtcacaga tacaattgtc gttactgaag gtgacggcat	660
ttcaacacca aaactcaaag aagactacca aattgggtgg tattctgagg ataggcactc	720
aggtgttaaa gactatgtcg ttgtacatgg ctatttcacc gaagtttact accagcttga	780



S226CAS111.ST25  
 gtctacacaa attactacag acactgggtat tgaaaatgct acattcttca tctttaacaa 840  
 gcttggttaa gaccaccga atgtgcaa acacacaatc gacggctctt caggagtgc 900  
 taatccagca atggatccaa tttatgatga gccgacgacg actactagcg tgcctttgta 960  
 agcacaagaa agtgagtacg aacttatgta ctcattcgtt tcggaagaaa cagggtacgtt 1020  
 aatagttaat agcgtacttc tttttcttgc tttcgtggta ttcttgctag tcacactagc 1080  
 catccttact gcgctt 1096

<210> 8<211> 1135<212> DNA<213> CORONAVIRUS<400> 8  
 attgccatcg tcatggttac aatcttgctt tgttgcatga ctagtgttg cagttgcctc 60  
 aaggggtgcat gctcttggtg ttcttgctgc aagtttgatg aggatgactc tgagccagtt 120  
 ctcaagggtg tcaaattaca ttacacataa acgaacttat ggatttggtt atgagatttt 180  
 ttactcttgg atcaattact gcacagccag taaaaattga caatgcttct cctgcaagta 240  
 ctgttcatgc tacagcaacg ataccgctac aagcctcact ccctttcgga tggcttggtta 300  
 ttggcggtgc atttcttgct gtttttcaga gcgctaccaa aataattgcg ctcaataaaa 360  
 gatggcagct agccctttat aagggcttcc agttcatttg caatttactg ctgctatttg 420  
 ttaccatcta ttcacatctt ttgcttgctg ctgcaggtat ggaggcgcaa tttttgtacc 480  
 tctatgcctt gatataattt ctacaatgca tcaacgcag tagaattatt atgagatgtt 540  
 ggctttgttg gaagtgcaaa tccaagaacc cattacttta tgatgccaac tactttgttt 600  
 gctggcacac acataactat gactactgta taccatataa cagtgtcaca gatacaattg 660  
 tcgttactga aggtgacggc atttcaacac caaaactcaa agaagactac caaattgggtg 720  
 gttattctga ggataggcac tcaggtgtta aagactatgt cgttgtagat ggctatttca 780  
 ccgaagtta ctaccagctt gactctacac aaattactac agacactggg attgaaaatg 840  
 ctacattctt catctttaac aagcttggtta aagaccacc gaatgtgcaa atacacacaa 900  
 tcgacggctc ttcaggagtt gctaattccag caatggatcc aatttatgat gagccgacga 960  
 cgactactag cgtgcctttg taagcacaag aaagttagta cgaacttatg tactcattcg 1020  
 tttcggaaga aacagggtacg ttaatagtta atagcgtact tctttttctt gctttcgtgg 1080  
 tattcttgct agtcacacta gccatcctta ctgcgcttcg attgtgtgcg tactg 1135

<210> 9<211> 1096<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (137)..(958)<223>

<400> 9  
 tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc tcttggtggtt 60  
 cttgctgcaa gtttgatgag gatgactctg agccagttct caagggtgtc aaattacatt 120  
 acacataaac gaactt atg gat ttg ttt atg aga ttt ttt act ctt gga tca 172  
                   Met Asp Leu Phe Met Arg Phe Phe Thr Leu Gly Ser  
                   1                  5                  10  
 att act gca cag cca gta aaa att gac aat gct tct cct gca agt act 220  
 Ile Thr Ala Gln Pro Val Lys Ile Asp Asn Ala Ser Pro Ala Ser Thr  
                   15                  20                  25

## S226CAS111.ST25

gtt cat gct aca gca acg ata ccg cta caa gcc tca ctc cct ttc gga Val His Ala Thr Ala Thr Ile Pro Leu Gln Ala Ser Leu Pro Phe Gly 30 35 40	268
tgg ctt gtt att ggc gtt gca ttt ctt gct gtt ttt cag agc gct acc Trp Leu Val Ile Gly Val Ala Phe Leu Ala Val Phe Gln Ser Ala Thr 45 50 55 60	316
aaa ata att gcg ctc aat aaa aga tgg cag cta gcc ctt tat aag ggc Lys Ile Ile Ala Leu Asn Lys Arg Trp Gln Leu Ala Leu Tyr Lys Gly 65 70 75	364
ttc cag ttc att tgc aat tta ctg ctg cta ttt gtt acc atc tat tca Phe Gln Phe Ile Cys Asn Leu Leu Leu Leu Phe Val Thr Ile Tyr Ser 80 85 90	412
cat ctt ttg ctt gtc gct gca ggt atg gag gcg caa ttt ttg tac ctc His Leu Leu Leu Val Ala Ala Gly Met Glu Ala Gln Phe Leu Tyr Leu 95 100 105	460
tat gcc ttg ata tat ttt cta caa tgc atc aac gca tgt aga att att Tyr Ala Leu Ile Tyr Phe Leu Gln Cys Ile Asn Ala Cys Arg Ile Ile 110 115 120	508
atg aga tgt tgg ctt tgt tgg aag tgc aaa tcc aag aac cca tta ctt Met Arg Cys Trp Leu Cys Trp Lys Cys Lys Ser Lys Asn Pro Leu Leu 125 130 135 140	556
tat gat gcc aac tac ttt gtt tgc tgg cac aca cat aac tat gac tac Tyr Asp Ala Asn Tyr Phe Val Cys Trp His Thr His Asn Tyr Asp Tyr 145 150 155	604
tgt ata cca tat aac agt gtc aca gat aca att gtc gtt act gaa ggt Cys Ile Pro Tyr Asn Ser Val Thr Asp Thr Ile Val Val Thr Glu Gly 160 165 170	652
gac ggc att tca aca cca aaa ctc aaa gaa gac tac caa att ggt ggt Asp Gly Ile Ser Thr Pro Lys Leu Lys Glu Asp Tyr Gln Ile Gly Gly 175 180 185	700
tat tct gag gat agg cac tca ggt gtt aaa gac tat gtc gtt gta cat Tyr Ser Glu Asp Arg His Ser Gly Val Lys Asp Tyr Val Val Val His 190 195 200	748
ggc tat ttc acc gaa gtt tac tac cag ctt gag tct aca caa att act Gly Tyr Phe Thr Glu Val Tyr Tyr Gln Leu Glu Ser Thr Gln Ile Thr 205 210 215 220	796
aca gac act ggt att gaa aat gct aca ttc ttc atc ttt aac aag ctt Thr Asp Thr Gly Ile Glu Asn Ala Thr Phe Phe Ile Phe Asn Lys Leu 225 230 235	844
gtt aaa gac cca ccg aat gtg caa ata cac aca atc gac ggc tct tca Val Lys Asp Pro Pro Asn Val Gln Ile His Thr Ile Asp Gly Ser Ser 240 245 250	892
gga gtt gct aat cca gca atg gat cca att tat gat gag ccg acg acg Gly Val Ala Asn Pro Ala Met Asp Pro Ile Tyr Asp Glu Pro Thr Thr 255 260 265	940
act act agc gtg cct ttg taagcacaag aaagtgagta cgaacttatg Thr Thr Ser Val Pro Leu 270	988
tactcattcg tttcggaaga aacaggtacg ttaatagtta atagcgtact tctttttctt	1048
gcttttcgtgg tattcttgct agtcacacta gccatcctta ctgcgctt	1096

&lt;210&gt; 10&lt;211&gt; 274&lt;212&gt; PRT&lt;213&gt; CORONAVIRUS&lt;400&gt; 10

Met Asp Leu Phe Met Arg Phe Phe Thr Leu Gly Ser Ile Thr Ala Gln  
1 5 10 15Pro Val Lys Ile Asp Asn Ala Ser Pro Ala Ser Thr Val His Ala Thr  
20 25 30Ala Thr Ile Pro Leu Gln Ala Ser Leu Pro Phe Gly Trp Leu Val Ile  
35 40 45Gly Val Ala Phe Leu Ala Val Phe Gln Ser Ala Thr Lys Ile Ile Ala  
50 55 60Leu Asn Lys Arg Trp Gln Leu Ala Leu Tyr Lys Gly Phe Gln Phe Ile  
65 70 75 80Cys Asn Leu Leu Leu Leu Phe Val Thr Ile Tyr Ser His Leu Leu Leu  
85 90 95Val Ala Ala Gly Met Glu Ala Gln Phe Leu Tyr Leu Tyr Ala Leu Ile  
100 105 110Tyr Phe Leu Gln Cys Ile Asn Ala Cys Arg Ile Ile Met Arg Cys Trp  
115 120 125Leu Cys Trp Lys Cys Lys Ser Lys Asn Pro Leu Leu Tyr Asp Ala Asn  
130 135 140Tyr Phe Val Cys Trp His Thr His Asn Tyr Asp Tyr Cys Ile Pro Tyr  
145 150 155 160Asn Ser Val Thr Asp Thr Ile Val Val Thr Glu Gly Asp Gly Ile Ser  
165 170 175Thr Pro Lys Leu Lys Glu Asp Tyr Gln Ile Gly Gly Tyr Ser Glu Asp  
180 185 190Arg His Ser Gly Val Lys Asp Tyr Val Val Val His Gly Tyr Phe Thr  
195 200 205Glu Val Tyr Tyr Gln Leu Glu Ser Thr Gln Ile Thr Thr Asp Thr Gly  
210 215 220Ile Glu Asn Ala Thr Phe Phe Ile Phe Asn Lys Leu Val Lys Asp Pro  
225 230 235 240Pro Asn Val Gln Ile His Thr Ile Asp Gly Ser Ser Gly Val Ala Asn  
245 250 255

S226CAS111.ST25

<210> 11<211> 1096<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
(558)..(1019)<223>

Page 33





S226CAS111.ST25  
150

140                      145                      150  
taatagttaa tagcgtactt ctttttcttg ctttcgtggt attcttgcta gtcacactag 1079  
ccatccttac tgcgctt 1096

<210> 12<211> 154<212> PRT<213> CORONAVIRUS<400> 12

Met Met Pro Thr Thr Leu Phe Ala Gly Thr His Ile Thr Met Thr Thr  
1                      5                      10                      15

Val Tyr His Ile Thr Val Ser Gln Ile Gln Leu Ser Leu Leu Lys Val  
20                      25                      30

Thr Ala Phe Gln His Gln Asn Ser Lys Lys Thr Thr Lys Leu Val Val  
35                      40                      45

Ile Leu Arg Ile Gly Thr Gln Val Leu Lys Thr Met Ser Leu Tyr Met  
50                      55                      60

Ala Ile Ser Pro Lys Phe Thr Thr Ser Leu Ser Leu His Lys Leu Leu  
65                      70                      75                      80

Gln Thr Leu Val Leu Lys Met Leu His Ser Ser Ser Leu Thr Ser Leu  
85                      90                      95

Leu Lys Thr His Arg Met Cys Lys Tyr Thr Gln Ser Thr Ala Leu Gln  
100                      105                      110

Glu Leu Leu Ile Gln Gln Trp Ile Gln Phe Met Met Ser Arg Arg Arg  
115                      120                      125

Leu Leu Ala Cys Leu Cys Lys His Lys Lys Val Ser Thr Asn Leu Cys  
130                      135                      140

Thr His Ser Phe Arg Lys Lys Gln Val Arg  
145                      150

<210> 13<211> 332<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
(36)..(263)<223>

<400> 13  
tgcctttgta agcacaagaa agtgagtacg aactt atg tac tca ttc gtt tcg 53  
Met Tyr Ser Phe Val Ser  
1                      5

gaa gaa aca ggt acg tta ata gtt aat agc gta ctt ctt ttt ctt gct 101  
Glu Glu Thr Gly Thr Leu Ile Val Asn Ser Val Leu Leu Phe Leu Ala  
10                      15                      20

ttc gtg gta ttc ttg cta gtc aca cta gcc atc ctt act gcg ctt cga 149  
Phe Val Val Phe Leu Leu Val Thr Leu Ala Ile Leu Thr Ala Leu Arg  
25                      30                      35

ttg tgt gcg tac tgc tgc aat att gtt aac gtg agt tta gta aaa cca 197  
Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn Val Ser Leu Val Lys Pro  
40                      45                      50

## S226CAS111.ST25

acg gtt tac gtc tac tcg cgt gtt aaa aat ctg aac tct tct gaa gga 245  
 Thr Val Tyr Val Tyr Ser Arg Val Lys Asn Leu Asn Ser Ser Glu Gly  
 55 60 65 70

gtt cct gat ctt ctg gtc taaacgaact aactattatt attattctgt 293  
 Val Pro Asp Leu Leu Val  
 75

ttggaacttt aacattgctt atcatggcag acaacggta 332

<210> 14<211> 76<212> PRT<213> CORONAVIRUS<400> 14  
 Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu Ile Val Asn Ser  
 1 5 10 15

Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala  
 20 25 30

Ile Leu Thr Ala Leu Arg Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn  
 35 40 45

Val Ser Leu Val Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn  
 50 55 60

Leu Asn Ser Ser Glu Gly Val Pro Asp Leu Leu Val  
 65 70 75

<210> 15<211> 332<212> DNA<213> CORONAVIRUS<400> 15 60  
 tgccttttgta agcacaagaa agtgagtacg aacttatgta ctcatcgtt tcggaagaaa  
 caggtagcgtt aatagttaat agcgtacttc tttttcttgc tttcgtggta ttcttgctag 120  
 tcacactagc catccttact gcgcttcgat tgtgtgcgta ctgctgcaat attgttaacg 180  
 tgagtttagt aaaaccaacg gtttacgtct actcgcgtgt taaaaatctg aactcttctg 240  
 aaggagtcc tgatcttctg gtctaaacga actaactatt attattattc tgtttggaac 300  
 ttttaacattg cttatcatgg cagacaacgg ta 332

<210> 16<211> 708<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (41)..(703)<223>

<400> 16 55  
 tattattatt attctgtttg gaactttaac attgcttatc atg gca gac aac ggt  
 Met Ala Asp Asn Gly  
 1 5

act att acc gtt gag gag ctt aaa caa ctc ctg gaa caa tgg aac cta 103  
 Thr Ile Thr Val Glu Glu Leu Lys Gln Leu Leu Glu Gln Trp Asn Leu  
 10 15 20

gta ata ggt ttc cta ttc cta gcc tgg att atg tta cta caa ttt gcc 151  
 Val Ile Gly Phe Leu Phe Leu Ala Trp Ile Met Leu Leu Gln Phe Ala  
 25 30 35

tat tct aat cgg aac agg ttt ttg tac ata ata aag ctt gtt ttc ctc 199  
 Tyr Ser Asn Arg Asn Arg Phe Leu Tyr Ile Ile Lys Leu Val Phe Leu  
 40 45 50

S226CAS111.ST25

tgg ctc ttg tgg cca gta aca ctt gct tgt ttt gtg ctt gct gct gtc 247  
 Trp Leu Leu Trp Pro Val Thr Leu Ala Cys Phe Val Leu Ala Ala Val  
 55 60 65

tac aga att aat tgg gtg act ggc ggg att gcg att gca atg gct tgt 295  
 Tyr Arg Ile Asn Trp Val Thr Gly Gly Ile Ala Ile Ala Met Ala Cys  
 70 75 80 85

att gta ggc ttg atg tgg ctt agc tac ttc gtt gct tcc ttc agg ctg 343  
 Ile Val Gly Leu Met Trp Leu Ser Tyr Phe Val Ala Ser Phe Arg Leu  
 90 95 100

ttt gct cgt acc cgc tca atg tgg tca ttc aac cca gaa aca aac att 391  
 Phe Ala Arg Thr Arg Ser Met Trp Ser Phe Asn Pro Glu Thr Asn Ile  
 105 110 115

ctt ctc aat gtg cct ctc cgg ggg aca att gtg acc aga ccg ctc atg 439  
 Leu Leu Asn Val Pro Leu Arg Gly Thr Ile Val Thr Arg Pro Leu Met  
 120 125 130

gaa agt gaa ctt gtc att ggt gct gtg atc att cgt ggt cac ttg cga 487  
 Glu Ser Glu Leu Val Ile Gly Ala Val Ile Ile Arg Gly His Leu Arg  
 135 140 145

atg gcc gga cac tcc cta ggg cgc tgt gac att aag gac ctg cca aaa 535  
 Met Ala Gly His Ser Leu Gly Arg Cys Asp Ile Lys Asp Leu Pro Lys  
 150 155 160 165

gag atc act gtg gct aca tca cga acg ctt tct tat tac aaa tta gga 583  
 Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser Tyr Tyr Lys Leu Gly  
 170 175 180

gcg tcg cag cgt gta ggc act gat tca ggt ttt gct gca tac aac cgc 631  
 Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe Ala Ala Tyr Asn Arg  
 185 190 195

tac cgt att gga aac tat aaa tta aat aca gac cac gcc ggt agc aac 679  
 Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp His Ala Gly Ser Asn  
 200 205 210

gac aat att gct ttg cta gta cag taagt 708  
 Asp Asn Ile Ala Leu Leu Val Gln  
 215 220

<210> 17<211> 221<212> PRT<213> CORONAVIRUS<400> 17

Met Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln Leu Leu  
 1 5 10 15

Glu Gln Trp Asn Leu Val Ile Gly Phe Leu Phe Leu Ala Trp Ile Met  
 20 25 30

Leu Leu Gln Phe Ala Tyr Ser Asn Arg Asn Arg Phe Leu Tyr Ile Ile  
 35 40 45

Lys Leu Val Phe Leu Trp Leu Leu Trp Pro Val Thr Leu Ala Cys Phe  
 50 55 60

Val Leu Ala Ala Val Tyr Arg Ile Asn Trp Val Thr Gly Gly Ile Ala  
 65 70 75 80

Ile Ala Met Ala Cys Ile Val Gly Leu Met Trp Leu Ser Tyr Phe Val

Ala Ser Phe Arg Leu Phe Ala Arg Thr Arg Ser Met Trp Ser Phe Asn  
100 105 110

Pro Glu Thr Asn Ile Leu Leu Asn Val Pro Leu Arg Gly Thr Ile Val  
115 120 125

Thr Arg Pro Leu Met Glu Ser Glu Leu Val Ile Gly Ala Val Ile Ile  
130 135 140

Arg Gly His Leu Arg Met Ala Gly His Ser Leu Gly Arg Cys Asp Ile  
145 150 155 160

Lys Asp Leu Pro Lys Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser  
165 170 175

Tyr Tyr Lys Leu Gly Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe  
180 185 190

Ala Ala Tyr Asn Arg Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp  
195 200 205

His Ala Gly Ser Asn Asp Asn Ile Ala Leu Leu Val Gln  
210 215 220

<210> 18<211> 769<212> DNA<213> CORONAVIRUS<400> 18  
cctgatcttc tgggtctaaac gaactaacta ttattattat tctgtttgga actttaacat 60  
tgcttatcat ggcagacaac ggtactatta ccgttgagga gcttaaaca ctcctggaac 120  
aatggaacct agtaataggt ttcctattcc tagcctggat tatgttacta caatttgcct 180  
attctaactg gaacagggtt ttgtacataa taaagcttgt tttcctctgg ctcttggtggc 240  
cagtaacact tgcttggttt gtgcttgctg ctgtctacag aattaattgg gtgactggcg 300  
ggattgcat tgcaatggct tgtattgtag gcttgatgtg gcttagctac ttcgttgctt 360  
ccttcaggct gtttgctcgt acccgctcaa tgtggtcatt caaccagaa acaaacattc 420  
ttctcaatgt gcctctccgg gggacaattg tgaccagacc gctcatggaa agtgaacttg 480  
tcattggtgc tgtgatcatt cgtggtcact tgcgaatggc cggacactcc ctagggcgct 540  
gtgacattaa ggacctgcca aaagagatca ctgtggctac atcacgaacg ctttcttatt 600  
acaaattagg agcgtcgcag cgtgtaggca ctgattcagg ttttgctgca tacaaccgct 660  
accgtattgg aaactataaa ttaaatacag accacgccgg tagcaacgac aatattgctt 720  
tgctagtaca gtaagtgaca acagatgttt catcttggtg acttccagg 769

<210> 19<211> 1231<212> DNA<213> CORONAVIRUS<400> 19  
taccgtattg gaaactataa attaaataca gaccacgccg gtagcaacga caatattgct 60  
ttgctagtac agtaagtgac aacagatgtt tcattctgtt gacttccagg ttacaatagc 120



S226CAS111.ST25

agagatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat	180
aataagttca atagttagac aattatttaa gcctctaact aagaagaatt attcggagtt	240
agatgatgaa gaacctatgg agtttagatta tccataaaac gaacatgaaa attattctct	300
tcctgacatt gattgtattt acatcttgcg agctatatca ctatcaggag tgtgttagag	360
gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat	420
ttcaccctct tgctgacaat aaatttgac taacttgac tagcacacac tttgcttttg	480
cttggtgctga cggtagctga catacctatc agctgcgtgc aagatcagtt tcaccaaacc	540
ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctcatgttg	600
ctgctctagt atttttaata ctttgcttca ccattaagag aaagacagaa tgaatgagct	660
cactttaatt gacttctatt tgtgcttttt agcctttctg ctattccttg ttttaataat	720
gcttattata ttttggtttt cactcgaaat ccaggatcta gaagaacctt gtaccaaagt	780
ctaaacgaac atgaaacttc tcattgtttt gacttgtatt tctctatgca gttgcatatg	840
cactgtagta cagcgctgtg catctaataa acctcatgtg cttgaagatc cttgtaaggt	900
acaacactag gggtaatact tatagcactg cttggctttg tgctctagga aagggtttac	960
cttttcatag atggcacact atggttcaaa catgcacacc taatgttact atcaactgtc	1020
aagatccagc tgggtggtgcg cttatagcta ggtgttggtta ctttcatgaa ggtcaccaaa	1080
ctgctgcatt tagagacgta cttgttgttt taaataaacg aacaaattaa aatgtctgat	1140
aatggacccc aatcaaacca acgtagtgcc ccccgatta catttggtgg acccacagat	1200
tcaactgaca ataaccagaa tggaggacgc a	1231

<210> 20<211> 1242<212> DNA<213> CORONAVIRUS<400> 20	
gcatacaacc gctaccgtat tggaaactat aaattaaata cagaccacgc cggtagcaac	60
gacaatattg ctttgctagt acagtaagtg acaacagatg tttcatcttg ttgacttcca	120
ggttacaata gcagagatat tgattatcat tatgaggact ttcaggattg ctatttgga	180
tcttgacgtt ataataagtt caatagttag acagttattt aagcctctaa ctaagaagaa	240
ttattcggag ttagatgatg aagaacctat ggagttagat tatccataaa acgaacatga	300
aaattattct cttcctgaca ttgattgtat ttacatcttg cgagctatat cactatcagg	360
agtgtgttag aggtacgact gtactactaa aagaaccttg cccatcagga acatacgagg	420
gcaattcacc atttcacct cttgctgaca ataaatttgc actaacttgc actagcacac	480
actttgcttt tgcttgctg gacggtagct gacataccta tcagctgcgt gcaagatcag	540
tttcacaaa acttttctac agacaagagg aggttcaaca agagctctac tcgccacttt	600
ttctcattgt tgctgctcta gtatttttaa tactttgctt caccattaag agaaagacag	660
aatgaatgag ctacttttaa ttgacttcta tttgtgcttt ttagcctttc tgctattcct	720
tgttttaata atgcttatta tattttggtt ttcactcgaa atccaggatc tagaagaacc	780
ttgtacaaa gtctaaacga acatgaaact tctcattgtt ttgacttgta tttctctatg	840

## S226CAS111.ST25

cagttgcata tgcactgtag tacagcgctg tgcattcaat aaacctcatg tgcttgaaga 900  
 tccttgtaag gtacaacact aggggtaata cttatagcac tgcttggctt tgtgctctag 960  
 gaaaggtttt accttttcat agatggcaca ctatggttca aacatgcaca cctaattgta 1020  
 ctatcaactg tcaagatcca gctgggtggtg cgcttatagc taggtgttgg taccttcatg 1080  
 aaggtcacca aactgctgca tttagagacg tacttggttgt tttaaataaa cgaacgaatt 1140  
 aaaatgtctg ataatggacc ccaatcaaac caacgtagtg cccccgcac tacatttggg 1200  
 ggaccacag attcaactga caataaccag aatggaggac gc 1242

<210> 21<211> 1231<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (86)..(274)<223>

<400> 21  
 taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60  
 ttgctagtac agtaagtac aacag atg ttt cat ctt gtt gac ttc cag gtt 112  
 Met Phe His Leu Val Asp Phe Gln Val  
 1 5  
 aca ata gca gag ata ttg att atc att atg agg act ttc agg att gct 160  
 Thr Ile Ala Glu Ile Leu Ile Ile Ile Met Arg Thr Phe Arg Ile Ala  
 10 15 20 25  
 att tgg aat ctt gac gtt ata ata agt tca ata gtg aga caa tta ttt 208  
 Ile Trp Asn Leu Asp Val Ile Ile Ser Ser Ile Val Arg Gln Leu Phe  
 30 35 40  
 aag cct cta act aag aag aat tat tcg gag tta gat gat gaa gaa cct 256  
 Lys Pro Leu Thr Lys Lys Asn Tyr Ser Glu Leu Asp Asp Glu Glu Pro  
 45 50 55  
 atg gag tta gat tat cca taaaacgaac atgaaaatta ttctcttcct 304  
 Met Glu Leu Asp Tyr Pro  
 60  
 gacattgatt gtattttacat cttgcgagct atatcactat caggagtgtg ttagaggtag 364  
 gactgtacta ctaaaagaac cttgcccatc aggaacatac gagggcaatt caccatttca 424  
 ccctcttgct gacaataaat ttgcactaac ttgcactagc acacactttg cttttgcttg 484  
 tgctgacggt actcgacata cctatcagct gcgtgcaaga tcagtttcac caaaactttt 544  
 catcagacaa gaggagggtc aacaagagct ctactcgcca cttttttctca ttgttgctgc 604  
 tctagtattt ttaatacttt gcttcacat taagagaaag acagaatgaa tgagctcact 664  
 ttaattgact tctatttggt ctttttagcc tttctgctat tccttggttt aataatgctt 724  
 atttatattt ggttttcact cgaaatccag gatctagaag aacctgtac caaagtctaa 784  
 acgaacatga aactttctcat tgttttgact tgtatttctc tatgcagttg catatgcact 844  
 gtagtacagc gctgtgcatc taataaacct catgtgcttg aagatccttg taaggtaaca 904  
 cactaggggt aatacttata gcactgcttg gctttgtgct ctaggaaagg ttttaccttt 964  
 tcatagatgg cacactatgg ttcaaacatg cacacctaatt gttactatca actgtcaaga 1024  
 tccagctggt ggtgcgctta tagctaggtg ttggtacctt catgaaggtc accaaactgc 1084  
 tgcatttaga gacgtacttg ttgttttaaa taaacgaaca aattaaaatg tctgataatg 1144

## S226CAS111.ST25

gacccaatc aaaccaacgt agtgcccccc gcattacatt tgggtggaccc acagattcaa 1204  
 ctgacaataa ccagaatgga ggacgca 1231

<210> 22<211> 63<212> PRT<213> CORONAVIRUS<400> 22

Met Phe His Leu Val Asp Phe Gln Val Thr Ile Ala Glu Ile Leu Ile  
 1 5 10 15

Ile Ile Met Arg Thr Phe Arg Ile Ala Ile Trp Asn Leu Asp Val Ile  
 20 25 30

Ile Ser Ser Ile Val Arg Gln Leu Phe Lys Pro Leu Thr Lys Lys Asn  
 35 40 45

Tyr Ser Glu Leu Asp Asp Glu Glu Pro Met Glu Leu Asp Tyr Pro  
 50 55 60

<210> 23<211> 1231<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (285)..(650)<223>

<400> 23  
 taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60  
 ttgctagtag agtaagtac aacagatggt tcattcttgg gacttccagg ttacaatagc 120  
 agagatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat 180  
 aataagttca atagttagac aattatttaa gcctctaact aagaagaatt attcggagtt 240  
 agatgatgaa gaacctatgg agtttagatta tccataaaac gaac atg aaa att att 296  
 Met Lys Ile Ile  
 1

ctc ttc ctg aca ttg att gta ttt aca tct tgc gag cta tat cac tat 344  
 Leu Phe Leu Thr Leu Ile Val Phe Thr Ser Cys Glu Leu Tyr His Tyr  
 5 10 15 20

cag gag tgt gtt aga ggt acg act gta cta cta aaa gaa cct tgc cca 392  
 Gln Glu Cys Val Arg Gly Thr Thr Val Leu Leu Lys Glu Pro Cys Pro  
 25 30 35

tca gga aca tac gag ggc aat tca cca ttt cac cct ctt gct gac aat 440  
 Ser Gly Thr Tyr Glu Gly Asn Ser Pro Phe His Pro Leu Ala Asp Asn  
 40 45 50

aaa ttt gca cta act tgc act agc aca cac ttt gct ttt gct tgt gct 488  
 Lys Phe Ala Leu Thr Cys Thr Ser Thr His Phe Ala Phe Ala Cys Ala  
 55 60 65

gac ggt act cga cat acc tat cag ctg cgt gca aga tca gtt tca cca 536  
 Asp Gly Thr Arg His Thr Tyr Gln Leu Arg Ala Arg Ser Val Ser Pro  
 70 75 80

aaa ctt ttc atc aga caa gag gag gtt caa caa gag ctc tac tcg cca 584  
 Lys Leu Phe Ile Arg Gln Glu Glu Val Gln Gln Glu Leu Tyr Ser Pro  
 85 90 95 100

ctt ttt ctc att gtt gct gct cta gta ttt tta ata ctt tgc ttc acc 632  
 Leu Phe Leu Ile Val Ala Ala Leu Val Phe Leu Ile Leu Cys Phe Thr  
 105 110 115

## S226CAS111.ST25

att aag aga aag aca gaa tgaatgagct cactttaatt gacttctatt 680  
 Ile Lys Arg Lys Thr Glu  
 120

tgtgcttttt agcctttctg ctattccttg ttttaataat gcttattata ttttggtttt 740  
 cactcgaaat ccaggatcta gaagaacctt gtaccaaagt ctaaacgaac atgaaacttc 800  
 tcattgtttt gacttgtatt tctctatgca gttgcatatg cactgtagta cagcgctgtg 860  
 catctaataa acctcatgtg cttgaagatc cttgtaaggt acaacactag gggtaatact 920  
 tatagcactg cttggctttg tgctctagga aaggttttac cttttcatag atggcacact 980  
 atggttcaaa catgcacacc taatgttact atcaactgtc aagatccagc tgggtggtgcg 1040  
 cttatagcta ggtgttggtta cttcatgaa ggtcaccaaa ctgctgcatt tagagacgta 1100  
 cttgttggtt taaataaacg aacaaattaa aatgtctgat aatggacccc aatcaaacca 1160  
 acgtagtgcc ccccgcata ctttgggtgg acccacagat tcaactgaca ataaccagaa 1220  
 tggaggacgc a 1231

<210> 24<211> 122<212> PRT<213> CORONAVIRUS<400> 24

Met Lys Ile Ile Leu Phe Leu Thr Leu Ile Val Phe Thr Ser Cys Glu  
 1 5 10 15

Leu Tyr His Tyr Gln Glu Cys Val Arg Gly Thr Thr Val Leu Leu Lys  
 20 25 30

Glu Pro Cys Pro Ser Gly Thr Tyr Glu Gly Asn Ser Pro Phe His Pro  
 35 40 45

Leu Ala Asp Asn Lys Phe Ala Leu Thr Cys Thr Ser Thr His Phe Ala  
 50 55 60

Phe Ala Cys Ala Asp Gly Thr Arg His Thr Tyr Gln Leu Arg Ala Arg  
 65 70 75 80

Ser Val Ser Pro Lys Leu Phe Ile Arg Gln Glu Glu Val Gln Gln Glu  
 85 90 95

Leu Tyr Ser Pro Leu Phe Leu Ile Val Ala Ala Leu Val Phe Leu Ile  
 100 105 110

Leu Cys Phe Thr Ile Lys Arg Lys Thr Glu  
 115 120

<210> 25<211> 1231<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
 (650)..(781)<223>

<400> 25  
 taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60  
 ttgctagtac agtaagtac aacagatgtt tcatcttggt gacttcagg ttacaatagc 120  
 agagatattg attatcatta tgaggacttt caggattgct atttggaaac ttgacgttat 180



## S226CAS111.ST25

aataagttca atagtgagac aattatttaa gcctctaact aagaagaatt attcggagtt 240  
agatgatgaa gaacctatgg agttagatta tccataaaac gaacatgaaa attattctct 300  
tcctgacatt gattgtatattt acatcttgcg agctatatca ctatcaggag tgtgttagag 360  
gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat 420  
ttcaccctct tgctgacaat aaatttgcac taacttgcac tagcacacac tttgcttttg 480  
cttgtgctga cggtagctga catacctatc agctgcgtgc aagatcagtt tcaccaaacc 540  
ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctcattgttg 600  
ctgctctagt atttttaata ctttgcttca ccattaagag aaagacaga atg aat gag 658  
Met Asn Glu  
1

ctc act tta att gac ttc tat ttg tgc ttt tta gcc ttt ctg cta ttc 706  
Leu Thr Leu Ile Asp Phe Tyr Leu Cys Phe Leu Ala Phe Leu Leu Phe  
5 10 15

ctt gtt tta ata atg ctt att ata ttt tgg ttt tca ctc gaa atc cag 754  
Leu Val Leu Ile Met Leu Ile Ile Phe Trp Phe Ser Leu Glu Ile Gln  
20 25 30 35

gat cta gaa gaa cct tgt acc aaa gtc taaacgaaca tgaaacttct 801  
Asp Leu Glu Glu Pro Cys Thr Lys Val  
40

cattgttttg acttgatattt ctctatgcag ttgcatatgc actgtagtac agcgctgtgc 861  
atctaataaaa cctcatgtgc ttgaagatcc ttgtaaggta caacactagg ggtaatactt 921  
atagcactgc ttggctttgt gctctaggaa aggtttttacc ttttcataga tggcacacta 981  
tggttcaaac atgcacacct aatgttacta tcaactgtca agatccagct ggtggtgcgc 1041  
ttatagctag gtgttggtac cttcatgaag gtcaccaaac tgctgcattt agagacgtac 1101  
ttgttgtttt aaataaacga acaaattaaa atgtctgata atggacccca atcaaaccaa 1161  
cgtagtgccc cccgcattac atttgggtgga cccacagatt caactgacaa taaccagaat 1221  
ggaggacgca 1231

<210> 26<211> 44<212> PRT<213> CORONAVIRUS<400> 26  
Met Asn Glu Leu Thr Leu Ile Asp Phe Tyr Leu Cys Phe Leu Ala Phe  
1 5 10 15  
Leu Leu Phe Leu Val Leu Ile Met Leu Ile Ile Phe Trp Phe Ser Leu  
20 25 30  
Glu Ile Gln Asp Leu Glu Glu Pro Cys Thr Lys Val  
35 40

<210> 27<211> 1231<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
(791)..(907)<223>  
<400> 27  
taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60

## S226CAS111.ST25

```

ttgctagtac agtaagtgac aacagatggt tcatcttggt gacttccagg ttacaatagc 120
agagatatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat 180
aataagttca atagtggagac aattatttaa gcctctaact aagaagaatt attcggagtt 240
agatgatgaa gaacctatgg agttagatta tccataaaac gaacatgaaa attattctct 300
tcctgacatt gattgtattt acatcttgcg agctatatca ctatcaggag tgtgttagag 360
gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat 420
ttcacctctt tgctgacaat aaatttgcac taacttgcac tagcacacac tttgcttttg 480
cttggtgctga cggctactga catacctatc agctgcgtgc aagatcagtt tcaccaaacc 540
ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctcatgttg 600
ctgctctagt atttttaata ctttgcctca ccattaagag aaagacagaa tgaatgagct 660
cactttaatt gacttctatt tgtgcttttt agcctttctg ctattccttg ttttaataat 720
gcttattata ttttggtttt cactcgaaat ccaggatcta gaagaacctt gtaccaaagt 780
ctaaacgaac atg aaa ctt ctc att gtt ttg act tgt att tct cta tgc 829
          1      5      10
          Met Lys Leu Leu Ile Val Leu Thr Cys Ile Ser Leu Cys

agt tgc ata tgc act gta gta cag cgc tgt gca tct aat aaa cct cat 877
Ser Cys Ile Cys Thr Val Val Gln Arg Cys Ala Ser Asn Lys Pro His
          15      20      25

gtg ctt gaa gat cct tgt aag gta caa cac taggggtaatt acttatagca 927
Val Leu Glu Asp Pro Cys Lys Val Gln His
          30      35

ctgcttggtt ttgtgctcta ggaaagggtt taccttttca tagatggcac actatggttc 987
aaacatgcac acctaattgtt actatcaact gtcaagatcc agctgggtgt gcgcttatag 1047
ctaggtgttg gtaccttcat gaaggtcacc aaactgctgc atttagagac gtacttggtg 1107
ttttaataaa acgaacaaat taaaatgtct gataatggac cccaatcaaa ccaacgtagt 1167
gccccccgca ttacatttgg tggaccacaca gattcaactg acaataacca gaatggagga 1227
cgca 1231

```

<210> 28<211> 39<212> PRT<213> CORONAVIRUS<400> 28

Met Lys Leu Leu Ile Val Leu Thr Cys Ile Ser Leu Cys Ser Cys Ile.  
1 5 10 15

Cys Thr Val Val Gln Arg Cys Ala Ser Asn Lys Pro His Val Leu Glu  
20 25 30

Asp Pro Cys Lys Val Gln His  
35

<210> 29<211> 1231<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
(876)..(1127)<223>

<400> 29  
taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60

Met Cys Leu Lys Ile Leu  
1 5

Val	Arg	Tyr	Asn	Thr	Arg	Gly	Asn	Thr	Ser	Thr	Ala	Trp	Leu	Cys
			10					15				20		

gct cta gga aag gtt tta cct ttt cat aga tgg cac act atg gtt caa  
Ala Leu Gly Lys Val Leu Pro Phe His Arg Trp His Thr Met Val Gln  
25 30 35

aca tgc aca cct aat gtt act atc aac tgt caa gat cca gct ggt ggt  
Thr Cys Thr Pro Asn Val Thr Ile Asn Cys Gln Asp Pro Ala Gly Gly  
40 45 50

gcg ctt ata gct agg tgt tgg tac ctt cat gaa ggt cac caa act gct  
Ala Leu Ile Ala Arg Cys Trp Tyr Leu His Glu Gly His Gln Thr Ala  
55 60 65 70

gca ttt aga gac gta ctt gtt gtt tta aat aaa cga aca aat  
Ala Phe Arg Asp Val Leu Val Val Leu Asn Lys Arg Thr Asn  
75 80

taaaatgtct gataatggac cccaatcaaa ccaacgtagt gcccccgca ttacatttgg  
tggacccaca gattcaactg acaataacca gaatggagga cgca

<210> 30<211> 84<212> PRT<213> CORONAVIRUS<400> 30

Met Cys Leu Lys Ile Leu Val Arg Tyr Asn Thr Arg Gly Asn Thr Tyr  
1 5 10 15

Ser Thr Ala Trp Leu Cys Ala Leu Gly Lys Val Leu Pro Phe His Arg  
20 25 30

Trp His Thr Met Val Gln Thr Cys Thr Pro Asn Val Thr Ile Asn Cys  
35 40 45

## S226CAS111.ST25

Gln Asp Pro Ala Gly Gly Ala Leu Ile Ala Arg Cys Trp Tyr Leu His  
 50 55 60

Glu Gly His Gln Thr Ala Ala Phe Arg Asp Val Leu Val Val Leu Asn  
 65 70 75 80

Lys Arg Thr Asn

```

<210> 31<211> 21221<212> DNA<213> CORONAVIRUS<400> 31
atggagagcc ttgttcttgg tgtcaacgag aaaacacacg tccaactcag ttgacctgtc 60
cttcagggtta gagacgtgct agtgcggtggc ttcgggggact ctgtggaaga ggccctatcg 120
gaggcacgtg aacacctcaa aaatggcact tgtgggtctag tagagctgga aaaaggcgta 180
ctgccccagc ttgaacagcc ctatgtgttc attaaacggt ctgatgcctt aagcaccaat 240
cacggccaca aggtcggtga gctggttgca gaaatggacg gcattcagta cggtcgtagc 300
ggataaacac tgggagtact cgtgccacat gtgggcgaaa cccaattgc ataccgcaat 360
gttcttcttc gtaagaacgg taataagga gccgggtggc atagctatgg catcgatcta 420
aagtcttatg acttaggtga cgagcttggc actgatccca ttgaagatta tgaacaaaac 480
tggaacacta agcatggcag tgggtgactc cgtgaactca ctcgtgagct caatggaggt 540
gcagtcactc gctatgtcga caacaatttc tgtggcccag atgggtaccc tcttgattgc 600
atcaaagatt ttctcgcacg cgcgggcaag tcaatgtgca ctctttccga acaacttgat 660
tacatcgagt cgaagagagg tgtctactgc tgccgtgacc atgagcatga aattgcctgg 720
ttactgagc gctctgataa gagctacgag caccagacac ccttcgaaat taagagtgcc 780
aagaaatttg acactttcaa aggggaatgc ccaaagtttg tgtttcctct taactcaaaa 840
gtcaaagtca ttcaaccacg tgttgaaaag aaaaagactg agggtttcat ggggcgtata 900
cgctctgtgt accctgttgc atctccacag gagtgtaaca atatgcactt gtctaccttg 960
atgaaatgta atcattgcga tgaagtttca tggcagacgt gcgactttct gaaagccact 1020
tgtgaacatt gtggcactga aaatttagtt attgaaggac ctactacatg tgggtaccta 1080
cctactaatg ctgtagtga aatgccatgt cctgcctgtc aagaccaga gattggacct 1140
gagcatagt ttgcagatta tcacaaccac tcaaacattg aaactcgact ccgcaagga 1200
ggtaggacta gatgttttgg aggctgtgtg tttgcctatg ttggctgcta taataagcgt 1260
gcctactggg ttccctcgtg tagtgctgat attggctcag gccatactgg cattactggt 1320
gacaatgtgg agaccttgaa tgaggatctc cttgagatac tgagtcgtga acgtgttaac 1380
attaacattg ttggcgattt tcatttgaat gaagaggttg ccatcatttt ggcatctttc 1440
tctgcttcta caagtgcctt tattgacact ataaagagtc ttgattacaa gtctttcaaa 1500
accattgttg agtcctgcgg taactataaa gttaccaagg gaaagcccgt aaaagggtgct 1560
tggaacattg gacaacagag atcagtttta acaccactgt gtggttttcc ctcacaggct 1620

```

## S226CAS111.ST25

gctggtgtta tcagatcaat ttttgcgcgc acacttgatg cagcaaacca ctcaattcct	1680
gatttgcaaa gagcagctgt caccatactt gatggtatct ctgaacagtc attacgtctt	1740
gtcgacgcca tggttttatac ttcagacctg ctcaccaaca gtgtcattat tatggcatat	1800
gtaactggtg gtcttgtaca acagacttct cagtggttgt ctaatctttt gggcactact	1860
gttgaaaaac tcaggcctat ctttgaatgg attgaggcga aacttagtgc aggagtggaa	1920
tttctcaagg atgcttgagg gattctcaaa tttctcatta cagggtgttt tgacatcgtc	1980
aagggtcaaa tacaggttgc ttcagataac atcaaggatt gtgtaaaatg cttcattgat	2040
gttggttaaca aggcactcga aatgtgcatt gatcaagtca ctatcgctgg cgcaaagttg	2100
cgatcactca acttaggtga agtcttcac gctcaaagca agggacttta ccgtcagtgt	2160
atacgtggca aggagcagct gcaactactc atgcctctta aggcaccaa agaagtaacc	2220
tttcttgaag gtgattcaca tgacacagta cttacctctg aggaggttgt tctcaagaac	2280
ggtgaactcg aagcactcga gacgcccgtt gatagcttca caaatggagc tatcgttggc	2340
acaccagtct gtgtaaatgg cctcatgctc ttagagatta aggacaaaga acaatactgc	2400
gcattgtctc ctggtttact ggctacaaac aatgtctttc gcttaaaagg ggggtgcacca	2460
attaaagggtg taacctttgg agaagatact gtttggaag ttcaagggtta caagaatgtg	2520
agaatcacat ttgagcttga tgaacgtgtt gacaaagtgc ttaatgaaaa gtgctctgtc	2580
tacactgttg aatccggtac cgaagttact gagtttgcac gtgtttagc agaggctgtt	2640
gtgaagactt tacaaccagt ttctgatctc cttaccaaca tgggtattga tcttgatgag	2700
tggagtgtag ctacattcta cttatttgat gatgctggtg aagaaaactt ttcacacgt	2760
atgtattgtt ctttttacc tccagatgag gaagaagagg acgatgcaga gtgtgaggaa	2820
gaagaaattg atgaaacctg tgaacatgag tacggtacag aggatgatta tcaaggtctc	2880
cctctggaat ttggtgcctc agctgaaaca gttcgagttg aggaagaaga agaggaagac	2940
tggctggatg atactactga gcaatcagag attgagccag aaccagaacc tacacctgaa	3000
gaaccagtta atcagtttac tggttattta aaacttactg acaatgttgc cattaaatgt	3060
gttgacatcg ttaaggaggc acaaagtgt aatcctatgg tgattgtaaa tgctgctaac	3120
atacacctga aacatggtgg tgggtgtagc ggtgcactca acaaggcaac caatggtgcc	3180
atgcaaaaagg agagtgtatg ttacattaag ctaaatggcc ctcttacagt aggagggtct	3240
tgtttgcttt ctggacataa tcttgctaag aagtgtctgc atgttggttg acctaaccta	3300
aatgcagggtg aggacatcca gcttcttaag gcagcatatg aaaatttcaa ttcacaggac	3360
atcttacttg caccattgtt gtcagcaggc atatttggtg ctaaaccact tcagtcttta	3420
caagtgtgcg tgcagacggt tcgtacacag gtttatattg cagtcaatga caaagctctt	3480
tatgagcagg ttgtcatgga ttatcttgat aacctgaagc ctagagtgga agcacctaaa	3540
caagaggagc caccaaacac agaagattcc aaaactgagg agaaatctgt cgtacagaag	3600
cctgtcgaatg tgaagccaaa aattaaggcc tgcattgatg aggttaccac aacactggaa	3660

## S226CAS111.ST25

gaaactaagt	ttcttaccaa	taagttactc	ttgtttgctg	atatcaatgg	taagctttac	3720
catgattctc	agaacatgct	tagaggtgaa	gatatgtctt	tccttgagaa	ggatgcacct	3780
tacatggtag	gtgatgttat	cactagtggt	gatatcactt	gtgttgtaat	accctccaaa	3840
aaggctggtg	gcactactga	gatgctctca	agagctttga	agaaaagtgcc	agttgatgag	3900
tatataacca	cgtaccctgg	acaaggatgt	gctgggtata	cacttgagga	agctaagact	3960
gctcttaaga	aatgcaaadc	tgcattttat	gtactacctt	cagaagcacc	taatgctaag	4020
gaagagattc	taggaactgt	atcctggaat	ttgagagaaa	tgcttgctca	tgctgaagag	4080
acaagaaaat	taatgcctat	atgcatggat	gttagagcca	taatggcaac	catccaacgt	4140
aagtataaag	gaattaaaat	tcaagagggc	atcggtgact	atgggtgtccg	attcttcttt	4200
tatactagta	aagagcctgt	agcttctatt	attacgaagc	tgaactctct	aaatgagccg	4260
cttgtcacia	tgccaattgg	ttatgtgaca	catggtttta	atcttgaaga	ggctgcgcgc	4320
tgtatgcgtt	ctcttaaagc	tcctgccgta	gtgtcagtat	catcaccaga	tgctgttact	4380
acataaatg	gatacctcac	ttcgtcatca	aagacatctg	aggagcactt	tgtagaaaca	4440
gtttcttttg	ctggctctta	cagagattgg	tcctattcag	gacagcgtac	agagttaggt	4500
gttgaatttc	ttaagcgtgg	tgacaaaatt	gtgtaccaca	ctctggagag	ccccgtcgag	4560
tttcatcttg	acggtgaggt	tctttcactt	gacaaactaa	agagtctctt	atccctgcgg	4620
gaggttaaga	ctataaaagt	gttcacaact	gtggacaaca	ctaactctca	cacacagctt	4680
gtggatatgt	ctatgacata	tggaacagag	tttgggtcaa	catacttgga	tggtgctgat	4740
gttacaaaaa	ttaaacttca	tgtaaatacat	gagggtaaga	ctttctttgt	actacctagt	4800
gatgacacac	tacgtagtga	agctttcgag	tactaccata	ctcttgatga	gagttttctt	4860
ggtaggtaca	tgtctgcttt	aaaccacaca	aagaaatgga	aatttcctca	agttggtggt	4920
ttaacttcaa	ttaaattggc	tgataacaat	tgttatttgt	ctagtgtttt	attagcactt	4980
caacagcttg	aagtcaaatt	caatgcacca	gcacttcaag	aggcttatta	tagagcccgt	5040
gctggtgatg	ctgctaactt	ttgtgcactc	atactcgctt	acagtaataa	aactgttggc	5100
gagcttggtg	atgtcagaga	aactatgacc	catcttctac	agcatgctaa	tttggaatct	5160
gcaaagcgag	ttcttaatat	ggtgtgtaaa	cattgtggtc	agaaaactac	taccttaacg	5220
ggtgtagaag	ctgtgatgta	tatgggtact	ctatcttatg	ataatcttaa	gacaggtggt	5280
tccattccat	gtgtgtgtgg	tcgtgatgct	acacaatatc	tagtacaaca	agagtcttct	5340
tttgttatga	tgtctgcacc	acctgctgag	tataaattac	agcaaggtag	attcttatgt	5400
gcgaatgagt	acactggtaa	ctatcagtgt	ggtcattaca	ctcatataac	tgctaaggag	5460
accctctatc	gtattgacgg	agctcacctt	acaaagatgt	cagagtacaa	aggaccagtg	5520
actgatgttt	tctacaagga	aacatcttac	actacaacca	tcaagcctgt	gtcgtataaa	5580
ctcgatggag	ttacttacac	agagattgaa	ccaaaattgg	atgggtatta	taaaaaggat	5640
aatgcttact	atacagagca	gcctatagac	cttgtaccaa	ctcaaccatt	accaaagtcg	5700

## S226CAS111.ST25

agttttgata atttcaaact cacatgttct aacacaaaat ttgctgatga tttaaatcaa 5760  
 atgacaggct tcacaaagcc agcttcacga gagctatctg tcacattctt cccagacttg 5820  
 aatggcgatg tagtggctat tgactataga cactattcag cgagtttcaa gaaagggtgct 5880  
 aaattactgc ataagccaat tgtttggcac attaaccagg ctacaaccaa gacaacgttc 5940  
 aaaccaaaaca cttggtgttt acgttgtctt tggagtacaa agccagtaga tacttcaaat 6000  
 tcatttgaag ttctggcagt agaagacaca caaggaatgg acaatcttgc ttgtgaaagt 6060  
 caacaaccca cctctgaaga agtagtgga aatcctacca tacagaagga agtcatagag 6120  
 tgtgacgtga aaactaccga agttgtaggc aatgtcatac ttaaaccatc agatgaagggt 6180  
 gttaaagtaa cacaagagtt aggtcatgag gatcttatgg ctgcttatgt ggaaaacaca 6240  
 agcattacca ttaagaaacc taatgagctt tcactagcct taggtttaaa aacaattgcc 6300  
 actcatggta ttgctgcaat taatagtgtt ccttggagta aaattttggc ttatgtcaaa 6360  
 ccattcttag gacaagcagc aattacaaca tcaaattgag ctaagagatt agcacaacgt 6420  
 gtgtttaaca atttatatgcc ttatgtgttt acattattgt tccaattgtg tacttttact 6480  
 aaaagtacca attctagaat tagagcttca ctacctacaa ctattgctaa aaatagtgtt 6540  
 aagagtgttg ctaaattatg tttggatgcc ggcattaatt atgtgaagtc acccaaattt 6600  
 tctaaattgt tcacaatcgc tatgtggcta ttgttgtaa gtatttgctt aggttctcta 6660  
 atctgtgtaa ctgctgcttt tgggtgactc ttatctaatt ttggtgctcc ttcttattgt 6720  
 aatggcggtta gagaattgta tcttaattcg tctaacgtta ctactatgga tttctgtgaa 6780  
 ggttcttttc cttgcagcat ttgtttaagt ggattagact cccttgattc ttatccagct 6840  
 cttgaaacca ttcagggtgac gatttcacg tacaagctag acttgacaat tttagggtctg 6900  
 gccgctgagt ggggttttggc atatatgttg ttcacaaaat tcttttattt attagggtctt 6960  
 tcagctataa tgcagggtgtt ctttggctat ttgctagtc atttcacag caattcttgg 7020  
 ctcagtgtgt ttatcattag tattgtacaa atggcacccg tttctgcaat ggtaggatg 7080  
 tacatcttct ttgcttctt ctactacata tggaagagct atgttcatat catggatggt 7140  
 tgcacctctt cgacttgcag gatgtgctat aagcgcaatc gtgccacacg cgttgagtgt 7200  
 acaactattg ttaatggcat gaagagatct ttctatgtct atgcaaatgg aggccgtggc 7260  
 ttctgcaaga ctcaaatg gaattgtctc aattgtgaca cattttgcac tggtagtaca 7320  
 ttcattagt atgaagttgc tcgtgatttg tcaactcagt ttaaaagacc aatcaaccct 7380  
 actgaccagt catcgtatat tgttgatagt gttgctgtga aaaatggcgc gcttcacctc 7440  
 tactttgaca aggctggtca aaagacctat gagagacatc cgctctcca tttgtcaat 7500  
 ttagacaatt tgagagctaa caacactaaa ggttcactgc ctattaatgt catagttttt 7560  
 gatggcaagt ccaaatgcga cgagtctgct tctaagtctg cttctgtgta ctacagtcag 7620  
 ctgatgtgcc aacctattct gttgcttgac caagctcttg tatcagacgt tggagatagt 7680  
 actgaagttt ccgttaagat gtttgatgct tatgtcgaca ctttttcagc aacttttagt 7740

## S226CAS111.ST25

gttcctatgg	aaaaacttaa	ggcacttggt	gctacagctc	acagcgagtt	agcaaagggg	7800
gtagcttttag	atgggtgtcct	ttctacattc	gtgtcagctg	cccgacaagg	tgttgttgat	7860
accgatgttg	acacaaagga	tgttattgaa	tgtctcaaac	tttcacatca	ctctgactta	7920
gaagtgacag	gtgacagttg	taacaatttc	atgctcacct	ataataaggt	tgaaaacatg	7980
acgcccagag	atcttggcgc	atgtattgac	tgtaatgcaa	ggcatatcaa	tgccaagta	8040
gcaaaaagtc	acaatgtttc	actcatctgg	aatgtaaaag	actacatgtc	tttatctgaa	8100
cagctgcgta	aacaaattcg	tagtgctgcc	aagaagaaca	acataccttt	tagactaact	8160
tgtgctacaa	ctagacaggt	tgtcaatgtc	ataactacta	aaatctcact	caagggtggt	8220
aagattgtta	gtacttgttt	taaacttatg	cttaaggcca	cattattgtg	cgttcttgct	8280
gcattgggtt	gttatatcgt	tatgccagta	catacattgt	caatccatga	tggttacaca	8340
aatgaaatca	ttgggttaca	agccattcag	gatgggtgtc	ctcgtgacat	catttctact	8400
gatgattgtt	ttgcaaataa	acatgctggt	tttgacgcat	ggtttagcca	gcgtggtggt	8460
tcatacaaaa	atgacaaaag	ctgccctgta	gtagctgcta	tcattacaag	agagattggt	8520
ttcatagtgc	ctggccttacc	gggtactgtg	ctgagagcaa	tcaatggtga	cttcttgcat	8580
tttctacctc	gtgttttttag	tgctgttggc	aacatttgct	acacaccttc	caaactcatt	8640
gagtatagtg	attttgctac	ctctgcttgc	gttcttgctg	ctgagtgtac	aatttttaag	8700
gatgctatgg	gcaaacctgt	gccatattgt	tatgacacta	at ttgctaga	gggttctatt	8760
tcttatagtg	agcttcgtcc	agacactcgt	tatgtgctta	tggatggttc	catcatacag	8820
tttcctaaca	cttacctgga	gggttctggt	agagtagtaa	caacttttga	tgctgagtac	8880
tgtagacatg	gtacatgcga	aaggtcagaa	gtaggtat tt	gcctatctac	cagtggtaga	8940
tgggttctta	ataatgagca	ttacagagct	ctatcaggag	ttttctgtgg	tgttgatgcg	9000
atgaatctca	tagctaacat	ctttactcct	cttggtgcaac	ctgtgggtgc	tttagatgtg	9060
tctgcttcag	tagtggctgg	tggtattatt	gccatattgg	tgacttggtc	tgccactact	9120
tttatgaaat	tcagacgtgt	ttttggtgag	tacaaccatg	ttgttgctgc	taatgcactt	9180
ttgtttttga	tgtctttcac	tatactctgt	ctggtaccag	cttacagctt	tctgccggga	9240
gtctactcag	tcttttactt	gtacttgaca	ttctatttca	ccaatgatgt	ttcattcttg	9300
gtcaccttc	aatggtttgc	catgttttct	cctattgtgc	ctttttggat	aacagcaatc	9360
tatgtattct	gtatttctct	gaagcactgc	cattggttct	ttaacaacta	tcttaggaaa	9420
agagtcatgt	ttaatggagt	tacatttagt	accttcgagg	aggctgcttt	gtgtaccttt	9480
ttgctcaaca	aggaaatgta	cctaaaattg	cgtagcgaga	cactgttgcc	acttacacag	9540
tataacaggt	atcttgctct	atataacaag	tacaagtatt	tcagtggagc	cttagatact	9600
accagctatc	gtgaagcagc	ttgctgccac	ttagcaaagg	ctctaaatga	ctttagcaac	9660
tcaggtgctg	atgttctcta	ccaaccacca	cagacatcaa	tcacttctgc	tgttctgcag	9720
agtggtttta	ggaaaatggc	attcccgtca	ggcaaagttg	aaggggtgcat	ggtacaagta	9780



## S226CAS111.ST25

acctgtggaa ctacaactct taatggattg tggttggatg acacagtata ctgtccaaga 9840  
 catgtcattt gcacagcaga agacatgctt aatcctaact atgaagatct gctcattcgc 9900  
 aaatccaacc atagcttttct tgttcaggct ggcaatgttc aacttcgtgt tattggccat 9960  
 tctatgcaaa attgtctgct taggcttaaa gttgatactt ctaaccctaa gacaccaag 10020  
 tataaatttg tccgtatcca acctgggtcaa acattttcag ttctagcatg ctacaatgggt 10080  
 tcaccatctg gtgtttatca gtgtgccatg agacctaatac ataccattaa aggttctttc 10140  
 cttaatggat catgtggtag tgttggtttt aacattgatt atgattgcgt gtctttctgc 10200  
 tatatgcatc atatggagct tccaacagga gtacacgctg gtactgactt agaaggtaaa 10260  
 ttctatgggtc ctttgttga cagacaaact gcacaggctg cagggtacaga cacaaccata 10320  
 acattaaatg ttttggcatg gctgtatgct gctgttatca atgggtgata gtggtttctt 10380  
 aatagattca ccactacttt gaatgacttt aaccttgttg caatgaagta caactatgaa 10440  
 cctttgacac aagatcatgt tgacatattg ggacctcttt ctgctcaaac aggaattgcc 10500  
 gtcttagata tgtgtgctgc tttgaaagag ctgctgcaga atgggtatgaa tggtcgtact 10560  
 atccttggtg gcactatttt agaagatgag ttacacccat ttgatgttgt tagacaatgc 10620  
 tctggtgtta ccttccaagg taagttcaag aaaattgtta agggcactca tcattggatg 10680  
 cttttaactt tcttgacatc actattgatt cttgttcaaa gtacacagtg gtcactgttt 10740  
 ttctttgttt acgagaatgc tttcttgcca ttactcttg gtattatggc aattgctgca 10800  
 tgtgctatgc tgcttgtaa gcataagcac gcattcttgt gcttgtttct gttaccttct 10860  
 cttgcaacag ttgcttactt taatatggtc tacatgcctg ctagctgggt gatgcgtatc 10920  
 atgacatggc ttgaattggc tgacactagc ttgtctggtt ataggcttaa ggattgtgtt 10980  
 atgtatgctt cagctttagt tttgcttatt ctcatgacag ctgcgactgt ttatgatgat 11040  
 gctgctagac gtgtttggac actgatgaat gtcattacac ttgtttacaa agtctactat 11100  
 ggtaatgctt tagatcaagc tatttccatg tgggccttag ttatttctgt aacctctaac 11160  
 tattctggtg tcgttacgac tatcatgttt ttagctagag ctatagtgtt tgtgtgtgtt 11220  
 gagtattacc cattgttatt tattactggc aacaccttac agtgtatcat gcttgtttat 11280  
 tgtttcttag gctattgttg ctgctgctac tttggccttt tctgtttact caaccgttac 11340  
 ttcaggctta ctcttggtgt ttatgactac ttggtctcta cacaagaatt taggtatatg 11400  
 aactcccagg ggcttttgcc tcctaagagt agtattgatg ctttcaagct taacattaag 11460  
 ttgttgggta ttggaggtaa accatgtatc aagggtgcta ctgtacagtc taaaatgtct 11520  
 gacgtaaagt gcacatctgt ggtactgctc tcggttcttc aacaacttag agtagagtca 11580  
 tcttctaaat tgtgggcaca atgtgtacaa ctccacaatg atattcttct tgcaaaagac 11640  
 acaactgaag ctttcgagaa gatgggttct cttttgtctg ttttgctatc catgcagggt 11700  
 gctgtagaca ttaataggtt gtgcgaggaa atgctcgata accgtgctac tcttcaggct 11760  
 attgcttcag aatttagttc tttaccatca tatgccgctt atgccactgc ccaggaggcc 11820

## S226CAS111.ST25

tatgagcagg	ctgtagctaa	tggtgattct	gaagtcgttc	tcaaaaagtt	aaagaaatct	11880
ttgaatgtgg	ctaaatctga	gtttgaccgt	gatgctgcca	tgcaacgcaa	gttggaaaag	11940
atggcagatc	aggctatgac	ccaaatgtac	aaacaggcaa	gatctgagga	caagagggca	12000
aaagtaacta	gtgctatgca	aacaatgctc	ttcactatgc	ttaggaagct	tgataatgat	12060
gcacttaaca	acattatcaa	caatgcgcgt	gatggttgtg	ttccactcaa	catcatacca	12120
ttgactacag	cagccaaact	catggttggt	gtccctgatt	atggtaccta	caagaacact	12180
tgtgatggta	acacctttac	atatgcatct	gcactctggg	aaatccagca	agttgttgat	12240
gcggatagca	agattgttca	acttagtgaa	attaacatgg	acaattcacc	aaatttggct	12300
tggcctctta	ttgttacagc	tctaagagcc	aactcagctg	ttaaactaca	gaataatgaa	12360
ctgagtccag	tagcactacg	acagatgtcc	tgtgcggctg	gtaccacaca	aacagcttgt	12420
actgatgaca	atgcacttgc	ctactataac	aattcgaagg	gaggtaggtt	tgtgctggca	12480
ttactatcag	accaccaaga	tctcaaattg	gctagattcc	ctaagagtga	tggtacaggt	12540
acaatttaca	cagaactgga	accaccttgt	aggtttgtta	cagacacacc	aaaagggcct	12600
aaagtgaat	acttgactt	catcaaaggc	ttaaacaacc	taaatagagg	tatggtgctg	12660
ggcagtttag	ctgctacagt	acgtcttcag	gctggaaatg	ctacagaagt	acctgccaat	12720
tcaactgtgc	tttccttctg	tgcttttgca	gtagaccctg	ctaaagcata	taaggattac	12780
ctagcaagtg	gaggacaacc	aatcaccaac	tgtgtgaaga	tgttgtgtac	acacactggg	12840
acaggacagg	caattactgt	aacaccagaa	gctaacatgg	accaagagtc	ctttgggtgg	12900
gcttcatggt	gtctgtattg	tagatgccac	attgaccatc	caaatcctaa	aggattctgt	12960
gacttgaaag	gtaagtacgt	ccaaatacct	accacttgtg	ctaattgacc	agtgggtttt	13020
acacttagaa	acacagtctg	taccgtctgc	ggaatgtgga	aaggttatgg	ctgtagttgt	13080
gaccaactcc	gcgaaccctt	gatgcagtct	gcggatgcat	caacgttttt	aaacgggttt	13140
gcggtgtaag	tgagcccgt	cttacaccgt	gcggcacagg	cactagtact	gatgtcgtct	13200
acagggcttt	tgatatattac	aacgaaaaag	ttgctgggtt	tgcaaagtcc	ctaaaaacta	13260
attgctgtcg	cttccaggag	aaggatgagg	aaggcaattt	attagactct	tactttgtag	13320
ttaagaggca	tactatgtct	aactaccaac	atgaagagac	tatttataac	ttggttaaag	13380
attgtccagc	ggttgctgtc	catgactttt	tcaagttag	agtagatgg	gacatggtac	13440
cacatatatc	acgtcagcgt	ctaactaaat	acacaatggc	tgatttagtc	tatgctctac	13500
gtcattttga	tgagggtaat	tgtgatacat	taaaagaaat	actcgtcaca	tacaattgct	13560
gtgatgatga	ttatttcaat	aagaaggatt	ggtatgactt	cgtagagaat	cctgacatct	13620
tacgcgtata	tgctaactta	ggtgagcgtg	tacgccaatc	attattaaag	actgtacaat	13680
tctgcgatgc	tatgcgtgat	gcaggcattg	taggcgtact	gacattagat	aatcaggatc	13740
ttaatgggaa	ctggtacgat	ttcgggtgatt	tcgtacaagt	agcaccaggc	tgccggagttc	13800
ctattgtgga	ttcatattac	tcattgctga	tgcccatcct	cactttgact	agggcattgg	13860

## S226CAS111.ST25

ctgctgagtc ccatatggat gctgatctcg caaaaccact tattaagtgg gatttgctga 13920  
 aatatgattt tacggaagag agactttgtc tcttcgaccg ttattttaaa tattgggacc 13980  
 agacatacca tcccaattgt attaactgtt tggatgatag gtgtatcctt cattgtgcaa 14040  
 actttaatgt gttattttct actgtgtttc cacctacaag ttttggacca ctagtaagaa 14100  
 aaatatttgt agatgggtgtt ccttttgttg tttcaactgg ataccatttt cgtgagttag 14160  
 gagtcgtaca taatcaggat gtaaacttac atagctcgcg tctcagtttc aaggaacttt 14220  
 tagtgatgc tgctgatcca gctatgcatg cagcttctgg caattttattg ctagataaac 14280  
 gcactacatg cttttcagta gctgcactaa caaacaatgt tgcttttcaa actgtcaaac 14340  
 ccggaatttt taataaagac ttttatgact ttgctgtgtc taaagggtttc tttaagggaag 14400  
 gaagtctgt tgaactaaaa cacttcttct ttgctcagga tggcaacgct gctatcagtg 14460  
 attatgacta ttatcgttat aatctgccaa caatgtgtga tatcagacaa ctctatttcg 14520  
 tagttgaagt tgttgataaa tactttgatt gttacgatgg tggctgtatt aatgccaacc 14580  
 aagtaatcgt taacaatctg gataaatcag ctgggttccc atttaataaa tggggtaagg 14640  
 ctagacttta ttatgactca atgagttatg aggatcaaga tgcacttttc gcgtatacta 14700  
 agcgtaatgt catccctact ataactcaaa tgaatcttaa gtatgccatt agtgcaaaga 14760  
 atagagctcg caccgtagct ggtgtctcta tctgtagtac tatgacaaat agacagtttc 14820  
 atcagaaatt attgaagtca atagccgcca ctagaggagc tactgtggta attggaacaa 14880  
 gcaagtttta cggtggctgg cataatatgt taaaaactgt ttacagtgat gtagaaaactc 14940  
 cacaccttat gggttgggat tatccaaaat gtgacagagc catgcctaac atgcttagga 15000  
 taatggcctc tcttgttctt gctcgcaaac ataacacttg ctgtaactta tcacaccggt 15060  
 tctacagggt agctaacgag tgtgcgcaag tattaagtga gatggtcatg tgtggcggct 15120  
 cactatatgt taaaccagggt ggaacatcat ccggtgatgc tacaactgct tatgctaata 15180  
 gtgtctttaa catttgctca gctgttacag ccaatgtaaa tgcacttctt tcaactgatg 15240  
 gtaataagat agctgacaag tatgtccgca atctacaaca caggctctat gagtgtctct 15300  
 atagaaatag ggatgttgat catgaattcg tggatgagtt ttacgcttac ctgcgtaaac 15360  
 atttctccat gatgattctt tctgatgatg ccgttgtgtg ctataacagt aactatgcgg 15420  
 ctcaagggtt agtagctagc attaagaact ttaaggcagt tctttattat caaaataatg 15480  
 tgttcatgtc tgaggcaaaa tgttggactg agactgacct tactaaagga cctcacgaat 15540  
 tttgtcaca gcatacaatg ctagttaaac aaggagatga ttacgtgtac ctgccttacc 15600  
 cagatccatc aagaatatta ggcgcaggct gttttgtcga tgatattgtc aaaacagatg 15660  
 gtacacttat gattgaaagg ttcgtgtcac tggctattga tgcttacca cttacaaaac 15720  
 atcctaataca ggagtatgct gatgtctttc acttgtattt acaatacatt agaaagttac 15780  
 atgatgagct tactggccac atgttggaca tgtattccgt aatgctaact aatgataaca 15840  
 cctcacggta ctgggaacct gagttttatg aggctatgta cacaccacat acagtcttgc 15900

## S226CAS111.ST25

aggctgtagg tgcttggtga ttgtgcaatt cacagacttc acttcgttgc ggtgcctgta 15960  
 ttaggagacc attcctatgt tgcaagtgt gctatgacca tgtcatttca acatcacaca 16020  
 aattagtgtt gtctgttaat ccctatgttt gcaatgcccc aggttgtgat gtcactgatg 16080  
 tgacacaact gtatctagga ggtatgagct attattgcaa gtcacataag cctcccatta 16140  
 gttttccatt atgtgctaata ggtcagggttt ttggtttata caaaaacaca tgtgtaggca 16200  
 gtgacaatgt cactgacttc aatgcgatag caacatgtga ttggactaat gctggcgatt 16260  
 acatacttgc caacacttgt actgagagac tcaagctttt cgcagcagaa acgctcaaag 16320  
 ccactgagga aacatttaag ctgtcatatg gtattgccac tgtacgcgaa gtactctctg 16380  
 acagagaatt gcatctttca tgggagggtg gaaaacctag accaccattg aacagaaact 16440  
 atgtctttac tggttaccgt gtaactaaaa atagtaaagt acagattgga gagtacacct 16500  
 ttgaaaaagg tgactatggt gatgctgttg tgtacagagg tactacgaca tacaagttga 16560  
 atgttggtga ttactttgtg ttgacatctc acactgtaat gccacttagt gcacctactc 16620  
 tagtgccaca agagcactat gtgagaatta ctggcttgta cccaacactc aacatctcag 16680  
 atgagttttc tagcaatgtt gcaaattatc aaaaggctcg catgcaaaaag tactctacac 16740  
 tccaaggacc acctggtact ggtaagagtc attttgccat cggacttgct ctctattacc 16800  
 catctgctcg catagtgtat acggcatgct ctcatgcagc tgttgatgcc ctatgtgaaa 16860  
 aggcatataa atatttgccc atagataaat gtagtagaat catacctgcg cgtgcgcgcg 16920  
 tagagtgttt tgataaattc aaagtgaatt caacactaga acagtatgtt ttctgactg 16980  
 taaatgcatt gccagaaaca actgctgaca ttgtagtctt tgatgaaatc tctatggcta 17040  
 ctaattatga cttgagtgtt gtcaatgcta gacttcgtgc aaaacactac gtctatattg 17100  
 gcgatcctgc tcaattacca gcccccgca cattgctgac taaaggcaca ctagaaccag 17160  
 aatattttta ttcagtgtgc agacttatga aaacaatagg tccagacatg ttccttgga 17220  
 cttgtcgccg ttgtcctgct gaaattgttg acactgtgag tgcttttagt tatgacaata 17280  
 agctaaaagc acacaaggat aagtcagctc aatgcttcaa aatgttctac aaagggtgta 17340  
 ttacacatga tgtttcatct gcaatcaaca gacctcaaat aggcgttgta agagaatttc 17400  
 ttacacgcaa tcctgcttgg agaaaagctg tttttatctc accttataat tcacagaacg 17460  
 ctgtagcttc aaaaatctta ggattgccta cgcagactgt tgattcatca cagggttctg 17520  
 aatatgacta tgtcatattc acacaaacta ctgaaacagc acactcttgt aatgtcaacc 17580  
 gcttcaatgt ggctatcaca agggcaaaaa ttggcatttt gtgcataatg tctgatagag 17640  
 atctttatga caaactgcaa ttacaagtc tagaaatacc acgtcgcaat gtggctacat 17700  
 tacaagcaga aaatgtaact ggacttttta aggactgtag taagatcatt actggctctc 17760  
 atcctacaca ggcacctaca cacctcagcg ttgatataaa gttcaagact gaaggattat 17820  
 gtgttgacat accaggcata ccaaggaca tgacctaccg tagactcatc tctatgatgg 17880  
 gtttcaaaat gaattaccaa gtcaatgggtt accctaatat gtttatcacc cgcgaagaag 17940

## S226CAS111.ST25

ctattcgtca cgttcgtgcg tggattggct ttgatgtaga gggctgtcat gcaactagag 18000  
 atgctgtggg tactaaccta cctctccagc taggattttc tacagggtgtt aacttagtag 18060  
 ctgtaccgac tggttatgtt gacactgaaa ataacacaga attcaccaga gttaatgcaa 18120  
 aacctccacc aggtgaccag tttaaacatc ttataccact catgtataaa ggcttgccct 18180  
 ggaatgtagt gcgtattaag atagtacaaa tgctcagtga tacactgaaa ggattgtcag 18240  
 acagagtcgt gttcgtcctt tgggcgcatg gctttgagct tacatcaatg aagtactttg 18300  
 tcaagattgg acctgaaaga acgtgttgct tgtgtgacaa acgtgcaact tgcttttcta 18360  
 cttcatcaga tacttatgcc tgctggaatc attctgtggg ttttgactat gtctataacc 18420  
 ctttatgat tgatgttcag cagtggggct ttacgggtaa cttcagagt aaccatgacc 18480  
 aacattgcca ggtacatgga aatgcacatg tggctagtgt tgatgctatc atgactagat 18540  
 gtttagcagt ccatgagtg cttgttaagc gcgttgattg gtctgttgaa taccctatta 18600  
 taggagatga actgagggtt aattctgctt gcagaaaagt acaacacatg gttgtgaagt 18660  
 ctgcattgct tgctgataag tttccagttc ttcattgacat tggaaatcca aaggctatca 18720  
 agtgtgtgcc tcaggctgaa gtagaatgga agttctacga tgctcagcca tgtagtga 18780  
 aagcttaca aatagaggaa ctcttctatt cttatgctac acatcacgat aaattcactg 18840  
 atggtgtttg tttgttttgg aattgtaacg ttgatcgta cccagccaat gcaattgtgt 18900  
 gtaggtttga cacaagagtc ttgtcaaact tgaacttacc aggctgtgat ggtggtagtt 18960  
 tgtatgtgaa taagcatgca ttccacactc cagctttcga taaaagtgca tttactaatt 19020  
 taaagcaatt gcctttcttt tactattctg atagtccttg tgagtctcat ggcaaacaag 19080  
 tagtgtcgga tattgattat gttccactca aatctgttac gtgtattaca cgatgcaatt 19140  
 taggtgggtg tgtttgcaga caccatgcaa atgagtaccg acagtacttg gatgcatata 19200  
 atatgatgat ttctgctgga tttagcctat ggatttaca acaatttgat acttataacc 19260  
 tgtggaatac atttaccagg ttacagagtt tagaaaatgt ggcttataat gttgttaata 19320  
 aaggacactt tgatggacac gccggcgaag cacctgtttc catcattaat aatgctgttt 19380  
 acacaaaggt agatggtatt gatgtggaga tctttgaaaa taagacaaca cttcctgtta 19440  
 atgttgcat ttagcttttg gctaagcgta acattaaacc agtgccagag attaagatac 19500  
 tcaataattt ggggtgtgat atcgctgcta atactgtaat ctgggactac aaaagagaag 19560  
 cccagcaca tgtatctaca ataggtgtct gcacaatgac tgacattgcc aagaaacct 19620  
 ctgagagtgc ttgttcttca cttactgtct tgtttgatgg tagagtggaa ggacaggtag 19680  
 acctttttag aaacgcccg aatgggtgtt taataacaga aggttcagtc aaaggctctaa 19740  
 caccttcaaa gggaccagca caagctagcg tcaatggagt cacattaatt ggagaatcag 19800  
 taaaaacaca gtttaactac tttaagaaag tagacggcat tattcaacag ttgcctgaaa 19860  
 cctactttac tcagagcaga gacttagagg attttaagcc cagatcaca atggaaactg 19920  
 actttctcga gctcgctatg gatgaattca tacagcgata taagctcgag ggctatgcct 19980

## S226CAS111.ST25

tcgaacacat cgtttatgga gatttcagtc atggacaact tggcgggtctt catttaatga 20040  
taggcttagc caagc̄gtca caagattcac cacttaaatt agaggatttt atccctatgg 20100  
acagcacagt gaaaaattac ttcataacag atgcgcaaac aggttcatca aaatgtgtgt 20160  
gttctgtgat tgatctttta cttgatgact ttgtcgagat aataaagtca caagatttgt 20220  
cagtgtattc aaaagtggc aaggttacaa ttgactatgc tgaaatttca ttcattgcttt 20280  
ggtgtaagga tggacatgtt gaaaccttct acccaaaact acaagcaagt caagcgtggc 20340  
aaccaggtgt tgcgatgcct aacttgatca agatgcaaag aatgcttctt gaaaagtgtg 20400  
accttcagaa ttatgggtgaa aatgctgtta taccaaaagg aataatgatg aatgtcgcaa 20460  
agtatactca actgtgtcaa tacttaata cacttacttt agctgtaccc tacaacatga 20520  
gagttattca ctttgggtgct ggctctgata aaggagtgtc accaggtaca gctgtgtctca 20580  
gacaatgggt gccaaactggc aactacttg tcgattcaga tcttaatgac ttcgtctccg 20640  
acgcagattc tactttaatt ggagactgtg caacagtaca tacggctaataaatgggacc 20700  
ttattattag cgatatgtat gaccctagga ccaaacatgt gacaaaagag aatgactcta 20760  
aagaagggtt tttcacttat ctgtgtggat ttataaagca aaaactagcc ctgggtgggtt 20820  
ctatagctgt aaagataaca gagcattctt ggaatgctga cctttacaag cttatgggcc 20880  
atttctcatg gtggacagct tttgttacaa atgtaaatgc atcatcatcg gaagcatttt 20940  
taattggggc taactatctt ggcaagccga aggaacaaat tgatggctat accatgcatg 21000  
ctaactacat tttctggagg aacacaaatc ctatccagtt gtcttcctat tcaactcttg 21060  
acatgagcaa atttctctt aaattaagag gaactgctgt aatgtctctt aaggagaatc 21120  
aaatcaatga tatgatttat tctcttctgg aaaaaggtag gcttatcatt agagaaaaca 21180  
acagagtgtt ggtttcaagt gatattcttg ttaacaacta a 21221

<210> 32<211> 297<212> DNA<213> CORONAVIRUS<400> 32  
atggaccca atcaaacca cgtagtgcc cccgcattac atttgggtgga cccacagatt 60  
caactgacaa taaccagaat ggaggacgca atggggcaag gccaaaacag cgccgacccc 120  
aaggtttacc caataatact gcgtcttggt tcacagctct cactcagcat ggcaaggagg 180  
aacttagatt ccctcgaggc cagggcggtc caatcaacac caatagtggc ccagatgacc 240  
aaattggcta ctaccgaaga gctacccgac gagttcgtgg tggtgacggc aaaatga 297

<210> 33<211> 98<212> PRT<213> CORONAVIRUS<400> 33

Met Asp Pro Asn Gln Thr Asn Val Val Pro Pro Ala Leu His Leu Val  
1 5 10 15

Asp Pro Gln Ile Gln Leu Thr Ile Thr Arg Met Glu Asp Ala Met Gly  
20 25 30

Gln Gly Gln Asn Ser Ala Asp Pro Lys Val Tyr Pro Ile Ile Leu Arg  
35 40 45

S226CAS111.ST25

Leu Gly Ser Gln Leu Ser Leu Ser Met Ala Arg Arg Asn Leu Asp Ser  
50 55 60

Leu Glu Ala Arg Ala Phe Gln Ser Thr Pro Ile Val Val Gln Met Thr  
65 70 75 80

Lys Leu Ala Thr Thr Glu Glu Leu Pro Asp Glu Phe Val Val Val Thr  
85 90 95

Ala Lys

<210> 34<211> 213<212> DNA<213> CORONAVIRUS<400> 34 60  
atgctgccac cgtgctacaa cttcctcaag gaacaacatt gccaaaaggc ttctacgcag  
aggggaagcag aggcggcagt caagcctctt ctcgctcctc atcacgtagt cgcggtaatt 120  
caagaaattc aactcctggc agcagtaggg gaaattctcc tgctcgaatg gctagcggag 180  
gtggtgaaac tgccctcgcg ctattgctgc tag 213

<210> 35<211> 70<212> PRT<213> CORONAVIRUS<400> 35  
Met Leu Pro Pro Cys Tyr Asn Phe Leu Lys Glu Gln His Cys Gln Lys  
1 5 10 15

Ala Ser Thr Gln Arg Glu Ala Glu Ala Ala Val Lys Pro Leu Leu Ala  
20 25 30

Pro His His Val Val Ala Val Ile Gln Glu Ile Gln Leu Leu Ala Ala  
35 40 45

Val Gly Glu Ile Leu Leu Leu Glu Trp Leu Ala Glu Val Val Lys Leu  
50 55 60

Pro Ser Arg Tyr Cys Cys  
65 70

<210> 36<211> 1377<212> DNA<213> CORONAVIRUS<220><221> CDS<222>  
(67)..(1335)<223>

<400> 36 60  
atgaaggcca ccaaactgct gcatttagag acgtacttgt tgttttaaataaacgaacaa  
attaaa atg tct gat aat gga ccc caa tca aac caa cgt agt gcc ccc 108  
Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro 1 5 10  
cgc att aca ttt ggt gga ccc aca gat tca act gac aat aac cag aat 156  
Arg Ile Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn 15 20 25 30  
gga gga cgc aat ggg gca agg cca aaa cag cgc cga ccc caa ggt tta 204  
Gly Gly Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu 35 40 45

## S226CAS111.ST25

ccc aat aat act gcg tct tgg ttc aca gct ctc act cag cat ggc aag Pro Asn Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys 50 55 60	252
gag gaa ctt aga ttc cct cga ggc cag ggc gtt cca atc aac acc aat Glu Glu Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn 65 70 75	300
agt ggt cca gat gac caa att ggc tac tac cga aga gct acc cga cga Ser Gly Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Arg Ala Thr Arg Arg 80 85 90	348
gtt cgt ggt ggt gac ggc aaa atg aaa gag ctc agc ccc aga tgg tac Val Arg Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr 95 100 105 110	396
ttc tat tac cta gga act ggc cca gaa gct tca ctt ccc tac ggc gct Phe Tyr Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala 115 120 125	444
aac aaa gaa ggc atc gta tgg gtt gca act gag gga gcc ttg aat aca Asn Lys Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr 130 135 140	492
ccc aaa gac cac att ggc acc cgc aat cct aat aac aat gct gcc acc Pro Lys Asp His Ile Gly Thr Arg Asn Pro Asn Asn Asn Ala Ala Thr 145 150 155	540
gtg cta caa ctt cct caa gga aca aca ttg cca aaa ggc ttc tac gca Val Leu Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala 160 165 170	588
gag gga agc aga ggc ggc agt caa gcc tct tct cgc tcc tca tca cgt Glu Gly Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg 175 180 185 190	636
agt cgc ggt aat tca aga aat tca act cct ggc agc agt agg gga aat Ser Arg Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn 195 200 205	684
tct cct gct cga atg gct agc gga ggt ggt gaa act gcc ctc gcg cta Ser Pro Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu 210 215 220	732
ttg ctg cta gac aga ttg aac cag ctt gag agc aaa gtt tct ggt aaa Leu Leu Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys 225 230 235	780
ggc caa caa caa caa ggc caa act gtc act aag aaa tct gct gct gag Gly Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu 240 245 250	828
gca tct aaa aag cct cgc caa aaa cgt act gcc aca aaa cag tac aac Ala Ser Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn 255 260 265 270	876
gtc act caa gca ttt ggg aga cgt ggt cca gaa caa acc caa gga aat Val Thr Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn 275 280 285	924
ttc ggg gac caa gac cta atc aga caa gga act gat tac aaa cat tgg Phe Gly Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp 290 295 300	972
ccg caa att gca caa ttt gct cca agt gcc tct gca ttc ttt gga atg Pro Gln Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met 305 310 315	1020



S226CAS111.ST25

tca cgc att ggc atg gaa gtc aca cct tcg gga aca tgg ctg act tat	1068
Ser Arg Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr	
320 325 330	
cat gga gcc att aaa ttg gat gac aaa gat cca caa ttc aaa gac aac	1116
His Gly Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn	
335 340 345 350	
gtc ata ctg ctg aac aag cac att gac gca tac aaa aca ttc cca cca	1164
Val Ile Leu Leu Asn Lys His Ile Asp Ala Tyr Lys Thr Phe Pro Pro	
355 360 365	
aca gag cct aaa aag gac aaa aag aaa aag act gat gaa gct cag cct	1212
Thr Glu Pro Lys Lys Asp Lys Lys Lys Lys Thr Asp Glu Ala Gln Pro	
370 375 380	
ttg ccg cag aga caa aag aag cag ccc act gtg act ctt ctt cct gcg	1260
Leu Pro Gln Arg Gln Lys Lys Gln Pro Thr Val Thr Leu Leu Pro Ala	
385 390 395	
gct gac atg gat gat ttc tcc aga caa ctt caa aat tcc atg agt gga	1308
Ala Asp Met Asp Asp Phe Ser Arg Gln Leu Gln Asn Ser Met Ser Gly	
400 405 410	
gct tct gct gat tca act cag gca taa acactcatga tgaccacaca	1355
Ala Ser Ala Asp Ser Thr Gln Ala	
415 420	
aggcagatgg gctatgtaaa cg	1377

<210> 37<211> 422<212> PRT<213> CORONAVIRUS<400> 37

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro Arg Ile  
1 5 10 15

Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn Gly Gly  
20 25 30

Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn  
35 40 45

Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Glu  
50 55 60

Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Gly  
65 70 75 80

Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Arg Ala Thr Arg Arg Val Arg  
85 90 95

Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr Phe Tyr  
100 105 110

Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala Asn Lys  
115 120 125

Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr Pro Lys  
130 135 140

S226CAS111.ST25

Asp His Ile Gly Thr Arg Asn Pro Asn Asn Asn Ala Ala Thr Val Leu  
145 150 155 160

Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly  
165 170 175

Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg Ser Arg  
180 185 190

Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn Ser Pro  
195 200 205

Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu Leu Leu  
210 215 220

Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys Gly Gln  
225 230 235 240

Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu Ala Ser  
245 250 255

Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn Val Thr  
260 265 270

Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly  
275 280 285

Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln  
290 295 300

Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met Ser Arg  
305 310 315 320

Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr His Gly  
325 330 335

Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn Val Ile  
340 345 350

Leu Leu Asn Lys His Ile Asp Ala Tyr Lys Thr Phe Pro Pro Thr Glu  
355 360 365

Pro Lys Lys Asp Lys Lys Lys Lys Thr Asp Glu Ala Gln Pro Leu Pro  
370 375 380

Gln Arg Gln Lys Lys Gln Pro Thr Val Thr Leu Leu Pro Ala Ala Asp  
385 390 395 400

Met Asp Asp Phe Ser Arg Gln Leu Gln Asn Ser Met Ser Gly Ala Ser  
405 410 415

## S226CAS111.ST25

Ala Asp Ser Thr Gln Ala  
420

```
<210> 38<211> 1377<212> DNA<213> CORONAVIRUS<400> 38
atgaaggtca ccaaactgct gcatttagag acgtacttgt tgttttaaat aaacgaacaa      60
attaaaaatgt ctgataatgg accccaatca aaccaacgta gtgccccccg cattacattt      120
ggtggaccca cagattcaac tgacaataac cagaatggag gacgcaatgg ggcaaggcca      180
aaacagcgcc gacccaaggg tttaaccaat aatactgcgt cttgggtcac agctctcact      240
cagcatggca aggaggaact tagattccct cgaggccagg gcgttccaat caacaccaat      300
agtgggccag atgaccaaat tggctactac cgaagagcta cccgacgagt tcgtgggtgg      360
gacggcaaaa tgaaagagct cagccccaga tgggtacttct attacctagg aactggccca      420
gaagcttcac ttccctacgg cgctaacaaa gaaggcatcg tatgggttgc aactgagggg      480
gccttgaata cacccaaaga ccacattggc acccgcaatc ctaataacaa tgctgccacc      540
gtgctacaac ttctcaagg aacaacattg ccaaaaggct tctacgcaga gggaagcaga      600
ggcggcagtc aagcctcttc tcgctcctca tcacgtagtc gcggtaatc aagaaattca      660
actcctggca gcagtagggg aaattctcct gctcgaatgg ctagcggagg tggtgaaact      720
gccctcgcg c tattgctgct agacagattg aaccagcttg agagcaaagt ttctggtaaa      780
ggccaacaac aacaaggcca aactgtcact aagaaatctg ctgctgaggc atctaaaaag      840
cctcgccaaa aacgtactgc caaaaacag tacaacgtca ctcaagcatt tgggagacgt      900
ggtccagaac aaaccaagg aaatttcggg gaccaagacc taatcagaca aggaactgat      960
tacaacatt ggccgcaa at tgcacaattt gctccaagt cctctgcatt ctttggaatg     1020
tcacgcattg gcatggaagt cacaccttcg ggaacatggc tgacttatca tggagccatt     1080
aaattggatg acaaagatcc acaattcaaa gacaacgtca tactgctgaa caagcacatt     1140
gacgcataca aaacattccc accaacagag ctaaaaaagg aaaaaagaa aaagactgat     1200
gaagctcagc ctttgccgca gagacaaaag aagcagccca ctgtgactct tcttcctgcg     1260
gctgacatgg atgatttctc cagacaactt caaaattcca tgagtggagc ttctgctgat     1320
tcaactcagg cataaacact catgatgacc acacaaggca gatgggctat gtaaacg      1377
```

```
<210> 39<211> 204<212> DNA<213> CORONAVIRUS<400> 39
atattaggtt ttacctacc caggaaaagc caaccaacct cgatctcttg tagatctgtt      60
ctctaaacga actttaaaat ctgtgtagct gtcgctcggc tgcatgccta gtgcacctac     120
gcagtataaa caataataaa ttttactgtc gttgacaaga aacgagtaac tcgtccctct     180
tctgcagact gcttacgggt tcgt      204
```

```
<210> 40<211> 809<212> DNA<213> CORONAVIRUS<400> 40
actcaagcat ttgggagacg tgggtccagaa caaacccaag gaaatttcgg ggaccaagac      60
ctaactcagac aaggaactga ttacaaacat tggccgcaaa ttgcacaatt tgctccaagt     120
```

S226CAS111.ST25

gcctctgcat tctttggaat gtcacgcatt ggcattggaag tcacaccttc gggaacatgg	180
ctgacttatac atggagccat taaattggat gacaaagatc cacaattcaa agacaacgtc	240
atactgctga acaagcacat tgacgcatac aaaacattcc caccaacaga gcctaaaaag	300
gacaaaaaga aaaagactga tgaagctcag cctttgccgc agagacaaaa gaagcagccc	360
actgtgactc ttcttcctgc ggctgacatg gatgatttct ccagacaact tcaaaattcc	420
atgagtggag cttctgctga ttcaactcag gcataaacac tcatgatgac cacacaaggc	480
agatgggcta tgtaaacgtt ttcgcaattc cgtttacgat acatagtcta ctcttgctga	540
gaatgaattc tcgtaactaa acagcacaaag taggtttagt taactttaat ctcacatagc	600
aatctttaat caatgtgtaa cattagggag gacttgaaag agccaccaca ttttcatcga	660
ggccacgcgg agtacgatcg agggtagagt gaataatgct agggagagct gcctatatgg	720
aagagcccta atgtgtaaaa ttaattttag tagtgctatc cccatgtgat ttaatatagct	780
tcttaggaga atgacaaaaa aaaaaaaaaa	809

<210> 41<211> 448<212> DNA<213> CORONAVIRUS<400> 41	
aatgaacaca tagggctgtt caagctgggg cagtacgcct tttccagct ctactagacc	60
acaagtgcc a tttttgaggt gttcacgtgc ctccgatagg gcctcttcca cagagtcccc	120
gaagccacgc actagcacgt ctctaacctg aaggacaggc aaactgagtt ggacgtgtgt	180
tttctcgttg acaccaagaa caaggctctc catcttacct ttcggtcaca cccggacgaa	240
acctaggtat gctgatgatc gactgcaaca cggacgaaac cgtaagcagt ctgcagaaga	300
gggacgagtt actcgtttct tgtcaacgac agtaaaattt attattgttt atactgcgta	360
ggtgcactag gcatgcagcc gagcgacagc tacacagatt ttaaagttcg tttagagaac	420
agatctacaa gagatcgagg ttggttgg	448

<210> 42<211> 2033<212> DNA<213> CORONAVIRUS<400> 42	
atacctaggt ttcgtccggg tgtgaccgaa aggtaagatg gagagccttg ttcttggtgt	60
caacgagaaa acacacgtcc aactcagttt gcctgtcctt caggttagag acgtgctagt	120
gcgtggcttc ggggactctg tggaagaggc cctatcgagg gcacgtgaac acctcaaaaa	180
tggcatttgt ggtctagtag agctggaaaa aggcgtactg cccagccttg aacagcccta	240
tgtgttcatt aaacgttctg atgccttaag caccaatcac ggccacaagg tcgttgagct	300
ggttgacgaa atggacggca ttcagtacgg tcgtagcggg ataactctgg gactactcgt	360
gccacatgtg ggcgaacccc caattgcata ccgcaatgtt cttcttcgta agaacggtaa	420
taagggagcc ggtggtcata gctatggcat cgatctaaag tcttatgact taggtgacga	480
gcttggcact gatccattg aagattatga acaaaactgg aacactaagc atggcagtgg	540
tgactccgt gaactcactc gtgagctcaa tggagggtgca gtcactcgct atgtcgacaa	600
caatttctgt ggcccagatg ggtaccctct tgattgcac aaagattttc tcgcacgcgc	660
gggcaagtca atgtgcactc tttccgaaca acttgattac atcgagtcga agagaggtgt	720

S226CAS111.ST25

ctactgctgc cgtgaccatg agcatgaaat tgcctgggtc actgagcgct ctgataagag	780
ctacgagcac cagacaccct tcgaaattaa gagtgccaag aaatttgaca ctttcaaagg	840
ggaatgcca aagtttgtgt ttctcttaa ctcaaagtc aaagtcattc aaccacgtgt	900
tgaaaagaaa aagactgagg gtttcatggg gcgtatacgc tctgtgtacc ctgttgcac	960
tccacaggag tgtaacaata tgcacttgtc taccttgatg aaatgtaatc attgcatga	1020
agtttcatgg cagacgtgcg actttctgaa agccacttgt gaacattgtg gcactgaaaa	1080
tttagttatt gaaggaccta ctacatgtgg gtacctacct actaatgctg tagtgaaaat	1140
gccatgtcct gcctgtcaag acccagagat tggacctgag catagtgttg cagattatca	1200
caaccactca aacattgaaa ctcgactccg caagggagggt aggactagat gttttggagg	1260
ctgtgtgttt gcctatgttg gctgctataa taagcgtgcc tactgggttc ctggtgctag	1320
tgctgatatt ggctcaggcc atactggcat tactggtgac aatgtggaga ccttgaatga	1380
ggatctcctt gagatactga gtcgtgaacg tgtaacatt aacattgttg gcgattttca	1440
tttgaatgaa gaggttgcca tcattttggc atctttctct gcttctacaa gtgcctttat	1500
tgacactata aagagtcttg attacaagtc tttaaaaacc attgttgagt cctgcggtaa	1560
ctataaagtt accaagggaa agcccgtaaa aggtgcttgg aacattggac aacagagatc	1620
agttttaaca ccactgtgtg gttttccctc acaggctgct ggtgttatca gatcaatttt	1680
tgcgcgacac cttgatgcag caaaccactc aattcctgat ttgcaaagag cagctgtcac	1740
catacttgat ggtatttctg aacagtcatt acgtcttgtc gacgccatgg ttatacttc	1800
agacctgctc accaacagtg tcattattat ggcataatga actggtggtc ttgtacaaca	1860
gacttctcag tggttgtcta atcttttggg cactactgtt gaaaaactca ggcctatctt	1920
tgaatggatt gaggcgaaac ttagtgcagg agttgaattt ctcaaggatg cttgggagat	1980
tctcaaattt ctcattacag gtgtttttga catcgtcaag ggtcaaatac agg	2033

<210> 43<211> 2018<212> DNA<213> CORONAVIRUS<400> 43	
ggattgaggc gaaacttagt gcaggagtgt aatttctcaa ggatgcttgg gagattctca	60
aatttctcat tacagggtgt tttgacatcg tcaagggtca aatacagggt gcttcagata	120
acatcaagga ttgtgtaaaa tgcttcattg atgttggtta caaggcactc gaaatgtgca	180
ttgatcaagt cactatcgct ggcgcaaagt tgcgatcact caacttaggt gaagtcttca	240
tcgctcaaag caagggactt taccgtcagt gtatacgtgg caaggagcag ctgcaactac	300
tcatgcctct taaggcacca aaagaagtaa cttttcttga aggtgattca catgacacag	360
tacttacctc tgaggagggt gttctcaaga acggtgaact cgaagcactc gagacgccc	420
ttgatagctt cacaaatgga gctatcggtg gcacaccagt ctgtgtaaat ggcctcatgc	480
tcttagagat taaggacaaa gaacaatact gcgcattgtc tcctgggttta ctggctacaa	540
acaatgtctt tcgcttaaaa gggggtgcac caattaaagg tgtaaccttt ggagaagata	600
ctgtttggga agttcaagggt tacaagaatg tgagaatcac atttgagctt gatgaacgtg	660

S226CAS111.ST25

ttgacaaagt gcttaatgaa aagtgtctctg tctacactgt tgaatccggt accgaagtta	720
ctgagtttgc atgtgtttgta gcagaggctg ttgtgaagac tttacaacca gtttctgac	780
tccttaccaa catgggtatt gatcttgatg agtggagtgt agctacattc tacttatttg	840
atgatgctgg tgaagaaaac ttttcatcac gtatgtattg ttccttttac cctccagatg	900
aggaagaaga ggacgatgca gagtgtgagg aagaagaaat tgatgaaacc tgtgaacatg	960
agtacggtac agaggatgat tatcaaggctc tccctctgga atttggtgcc tcagctgaaa	1020
cagttcgagt tgaggaagaa gaagaggaag actggctgga tgatactact gagcaatcag	1080
agattgagcc agaaccagaa cctacacctg aagaaccagt taatcagttt actggttatt	1140
taaaacttac tgacaatggt gccattaaat gtgttgacat cgtaaggag gcacaaagt	1200
ctaatacctat ggtgattgta aatgctgcta acatacacct gaaacatggt ggtggtgtag	1260
caggtgcact caacaaggca accaatggtg ccatgcaaaa ggagagtgat gattacatta	1320
agctaaatgg cctctttaca gtaggagggt cttgtttgct ttctggacat aatcttgcta	1380
agaagtgtct gcatgttggt ggacctaac taaatgcagg tgaggacatc cagcttctta	1440
aggcagcata tgaaaatttc aattcacagg acatcttact tgcaccattg ttgtcagcag	1500
gcatatttg tgctaaacca cttcagtctt tacaagtgtg cgtgcagacg gttcgtacac	1560
aggtttatat tgcagtcaat gacaaagctc tttatgagca ggtgtcatg gattatcttg	1620
ataacctgaa gcctagagtg gaagcaccta aacaagagga gccaccaaac acagaagatt	1680
ccaaaactga ggagaaatct gtcgtacaga agcctgtcga tgtgaagcca aaaattaagg	1740
cctgcattga tgaggttacc acaactggtg aagaaactaa gtttcttacc aataagttac	1800
tcttgtttgc tgatatcaat ggtaagcttt accatgattc tcagaacatg cttagagggtg	1860
aagatatgtc tttccttgag aaggatgcac cttacatggt aggtgatgtt atcactagt	1920
gtgatatcac ttgtgttgta ataccctcca aaaaggctgg tggcactact gagatgctct	1980
caagagcttt gaagaaagtg ccagttgatg agtatata	2018

<210> 44<211> 1442<212> DNA<213> CORONAVIRUS<400> 44

ttgatgaggt taccacaaca ctggaagaaa ctaagtttct taccaataag ttactcttgt	60
ttgctgatat caatggtaag ctttaccatg attctcagaa catgcttaga ggtgaagata	120
tgtcttttct tgagaaggat gcaccttaca tggtaggtga tggtatcact agtggtgata	180
tcacttggtg tgtaataccc tccaaaaagg ctggtggcac tactgagatg ctctcaagag	240
ctttgaagaa agtgccagtt gatgagtata taaccacgta ccctggacaa ggatgtgctg	300
gttatacact tgaggaagct aagactgctc ttaagaaatg caaatctgca ttttatgtac	360
taccttcaga agcacctaata gctaaggaag agattctagg aactgtatcc tggaatttga	420
gagaaatgct tgctcatgct gaagagacaa gaaaattaat gcctatatgc atggatgtta	480
gagccataat ggcaaccatc caacgtaagt ataaaggaat taaaattcaa gagggcatcg	540
ttgactatgg tgtccgattc ttcttttata ctagtaaaga gcctgtagct tctattatta	600

S226CAS111.ST25

cgaagctgaa ctctctaaat gagccgcttg tcacaatgcc aattgggttat gtgacacatg	660
gttttaaatct tgaagaggct gcgcgctgta tgcgttctct taaagctcct gccgtagtgt	720
cagtatcatc accagatgct gttactacat ataatggata cctcacttcg tcatcaaaga	780
catctgagga gcactttgta gaaacagttt ctttggtgctg ctcttacaga gattgggtcct	840
attcaggaca gcgtacagag ttaggtgttg aatttcttaa gcgtgggtgac aaaattgtgt	900
accacactct ggagagcccc gtcgagtttc atcttgacgg tgagggttctt tcacttgaca	960
aactaaagag tctcttatcc ctgcgggagg ttaagactat aaaagtgttc acaactgtgg	1020
acaacactaa tctccacaca cagcttggtg atatgtctat gacatatgga cagcagtttg	1080
gtccaacata cttggatggg gctgatgtta caaaaattaa acctcatgta aatcatgagg	1140
gtaagacttt ctttgtagta cctagtgtg acacactacg tagtgaagct ttcgagtact	1200
accatactct tgatgagagt tttcttggtg ggtacatgtc tgctttaaac cacacaaaga	1260
aatggaaatt tcctcaagtt ggtgggttaa cttcaattaa atgggctgat aacaattgtt	1320
atttgtctag tgttttatta gcacttcaac agcttgaagt caaattcaat gcaccagcac	1380
ttcaagaggc ttattataga gcccgtgctg gtgatgctgc taacttttgt gcactcatac	1440
tc	1442

<210> 45<211> 1050<212> DNA<213> CORONAVIRUS<400> 45	
atatgtctat gacatatgga cagcagtttg gtccaacata cttggatggg gctgatgtta	60
caaaaattaa acctcatgta aatcatgagg gtaagacttt ctttgtagta cctagtgtg	120
acacactacg tagtgaagct ttcgagtact accatactct tgatgagagt tttcttggtg	180
ggtacatgtc tgctttaaac cacacaaaga aatggaaatt tcctcaagtt ggtgggttaa	240
cttcaattaa atgggctgat aacaattgtt atttgtctag tgttttatta gcacttcaac	300
agcttgaagt caaattcaat gcaccagcac ttcaagaggc ttattataga gcccgtgctg	360
gtgatgctgc taacttttgt gcactcatac tcgcttacag taataaaact gttggcgagc	420
ttggtgatgt cagagaaact atgaccatc ttctacagca tgctaatttg gaatctgcaa	480
agcgagttct taatgtgggtg tgtaaacatt gtgggtcagaa aactactacc ttaacgggtg	540
tagaagctgt gatgtatatg ggtactctat cttatgataa tcttaagaca ggtgtttcca	600
ttccatgtgt gtgtggctgt gatgctacac aatatctagt acaacaagag tcttcttttg	660
ttatgatgtc tgcaccacct gctgagtata aattacagca aggtacattc ttatgtgcga	720
atgagtacac tggttaactat cagtgtggtc attacactca tataactgct aaggagaccc	780
tctatcgtat tgacggagct caccttaca agatgtcaga gtacaaagga ccagtgactg	840
atgttttcta caaggaaaca tcttacacta caaccatcaa gcctgtgtcg tataaactcg	900
atggagttac ttacacagag attgaaccaa aattggatgg gtattataaa aaggataatg	960
cttactatac agagcagcct atagaccttg taccaactca accattacca aatgcgagtt	1020
ttgataattt caaactcaca tgttctaaca	1050

## S226CAS111.ST25

<210> 46<211> 1995<212> DNA<213> CORONAVIRUS<400> 46  
 tttgtgcact catactcgct tacagtaata aaactgttgg cgagcttggt gatgtcagag 60  
 aaactatgac ccattcttcta cagcatgcta atttggaatc tgcaaagcga gttcttaatg 120  
 tgggtgtgtaa acattgtggg cagaaaacta ctaccttaac ggggtgtaaa gctgtgatgt 180  
 atatgggtac tctatcttat gataatctta agacaggtgt ttccattcca tgtgtgtgtg 240  
 gtcgtgatgc tacacaatat ctagtacaac aagagtcctt ttttgttatg atgtctgcac 300  
 cacctgctga gtataaatta cagcaaggta cattcttatg tgcgaatgag tacactggta 360  
 actatcagtg tggtcattac actcatataa ctgctaagga gacctcttat cgtattgacg 420  
 gagctcacct tacaagatg tcagagtaca aaggaccagt gactgatgtt ttctacaagg 480  
 aaacatctta cactacaacc atcaagcctg tgtcgtataa actcgatgga gttacttaca 540  
 cagagattga accaaaattg gatgggtatt ataaaaagga taatgcttac tatacagagc 600  
 agcctataga ccttgtacca actcaaccat taccaaatgc gagttttgat aatttcaaac 660  
 tcacatgttc taacacaaaa tttgctgatg atttaaataa aatgacaggc ttcacaaagc 720  
 cagcttcacg agagctatct gtcacattct tcccagactt gaatggcgat gtagtggcta 780  
 ttgactatag acactattca gcgagtttca agaaagggtg taaattactg cataagccaa 840  
 ttgtttggca cattaaccag gctacaacca agacaacgtt caaaccaaac acttggtgtt 900  
 tacgttgtct ttggagtaca aagccagtag atacttcaaa ttcatttgaa gttctggcag 960  
 tagaagacac acaaggaatg gacaatcttg cttgtgaaag tcaacaacc acctctgaag 1020  
 aagtagtgga aaatcctacc atacagaagg aagtcataga gtgtgacgtg aaaactaccg 1080  
 aagttgtagg caatgtcata cttaaaccat cagatgaagg tgttaaagta acacaagagt 1140  
 taggtcatga ggatcttatg gctgcttatg tggaaaacac aagcattacc attaagaaac 1200  
 ctaatgagct ttcactagcc ttaggtttta aaacaattgc cactcatggt attgctgcaa 1260  
 ttaatagtgt tccttggagt aaaattttgg cttatgtcaa accattctta ggacaagcag 1320  
 caattacaac atcaaattgc gctaagagat tagcacaacg tgtgtttaac aattatatgc 1380  
 cttatgtgtt tacattattg ttccaattgt gtacttttac taaaagtacc aattctagaa 1440  
 ttagagcttc actacctaca actattgcta aaaatagtgt taagagtgtt gctaaattat 1500  
 gtttgatgc cggcattaat tatgtgaagt cacccaaatt ttctaaattg ttcacaatcg 1560  
 ctatgtggct attgttgta agtatttgct taggttctct aatctgtgta actgctgctt 1620  
 ttggtgtact cttatctaatt tttggtgctc cttcttattg taatggcggt agagaattgt 1680  
 atcttaattc gtctaacgtt actactatgg atttctgtga aggttctttt ccttgacagca 1740  
 tttgtttaag tggattagac tcccttgatt cttatccagc tcttgaaacc attcaggtga 1800  
 cgatttcacg gtacaagcta gacttgacaa ttttaggtct ggccgctgag tgggttttgg 1860  
 catatatgtt gttcacaaaa ttcttttatt tattaggtct ttcagctata atgcaggtgt 1920  
 tctttggcta ttttgctagt catttcatca gcaattcttg gctcatgtgg tttatcatta 1980



gtattgtaca aatgg

1995

<210> 47<211> 1884<212> DNA<213> CORONAVIRUS<400> 47  
 aattcttggc tcatgtggtt tatcattagt attgtacaaa tggcaccggt ttctgcaatg 60  
 gttaggatgt acatcttctt tgcttctttc tactacatat ggaagagcta tgttcatatc 120  
 atggatgggt gcacctcttc gacttgcatt atgtgctata agcgcaatcg tgccacacgc 180  
 gttgagtgt caactattgt taatggcatg aagagatctt tctatgtcta tgcaaattgga 240  
 ggccgtggct tctgcaagac tcacaattgg aattgtctca attgtgacac attttgcact 300  
 ggtagtacat tcattagtga tgaagttgct cgtgatttgt cactccagtt taaaagacca 360  
 atcaacccta ctgaccagtc atcgtatat gttgatagtg ttgctgtgaa aaatggcgcg 420  
 cttcacctct actttgacaa ggctgggtcaa aagacctatg agagacatcc gctctcccat 480  
 tttgtcaatt tagacaattt gagagctaac aacactaaag gttcactgcc tattaatgtc 540  
 atagtttttg atggcaagtc caaatgagac gagtctgctt ctaagtctgc ttctgtgtac 600  
 tacagtcagc tgatgtgcca acctattctg ttgcttgacc aagctcttgt atcagacggt 660  
 ggagatagta ctgaagtttc cgttaagatg tttgatgctt atgtcgacac cttttcagca 720  
 acttttagtg ttcctatgga aaaacttaag gcacttggtg ctacagctca cagcgagtta 780  
 gcaaagggtg tagctttaga tgggtgtcctt tctacattcg tgtcagctgc cgcacaagggt 840  
 gttgttgata ccgatgttga cacaaggat gttattgaat gtctcaaaact ttcacatcac 900  
 tctgacttag aagtgcaggg tgacagttgt aacaatttca tgctcaccta taataagggt 960  
 gaaaacatga cgcccagaga tcttggcgca tgtattgact gtaatgcaag gcatatcaat 1020  
 gcccaagtag caaaaagtca caatgtttca ctcatctgga atgtaaaaga ctacatgtct 1080  
 ttatctgaac agctgcgtta acaaattcgt agtgcgtgcca agaagaacaa catacctttt 1140  
 agactaactt gtgctacaac tagacagggt gtcaatgtca taactactaa aatctcactc 1200  
 aagggtggta agattgttag tacttggttt aaacttatgc ttaaggccac attattgtgc 1260  
 gttcttgctg cattgggttg ttatatcggt atgccagtag atacattgtc aatccatgat 1320  
 ggttacacaa atgaaatcat tggttacaaa gccattcagg atgggtgtcac tcgtgacatc 1380  
 atttctactg atgattgttt tgcaataaaa catgctgggt ttgacgcatg gtttagccag 1440  
 cgtgggtggt catacaaaaa tgacaaaagc tgccctgtag tagctgctat cattacaaga 1500  
 gagattgggt tcatagtgcc tggcttaccg ggtactgtgc tgagagcaat caatgggtgac 1560  
 ttcttgcatt ttctacctcg tgtttttagt gctgttggca acatttgcta cacaccttcc 1620  
 aaactcattg agtatagtga ttttgcctac tctgcttgcg ttcttgctgc tgagtgtaca 1680  
 atttttaagg atgctatggg caaacctgtg ccatattgtt atgacactaa tttgctagag 1740  
 ggttctatct cttatagtga gcttcgtcca gacactcgtt atgtgcttat ggatgggttcc 1800  
 atcatacagt ttcctaacac ttacctggag ggttctgtta gagtagtaac aacttttgat 1860  
 gctgagtact gtagacatgg taca 1884

## S226CAS111.ST25

```

<210> 48<211> 2020<212> DNA<213> CORONAVIRUS<400> 48
cactcgttat gtgcttatgg atggttccat catacagttt cctaacactt acctggaggg 60
ttctgttaga gtagtaacaa cttttgatgc tgagtactgt agacatggta catgcgaaag 120
gtcagaagta ggtatttgcc tatctaccag tggtagatgg gttcttaata atgagcatta 180
cagagctcta tcaggagttt tctgtggtgt tgatgcgatg aatctcatag ctaacatctt 240
tactcctctt gtgcaacctg tgggtgcttt agatgtgtct gcttcagtag tggctggtgg 300
tattattgcc atattggtga cttgtgctgc ctactacttt atgaaattca gacgtgtttt 360
tggtgagtac aaccatgttg ttgctgctaa tgcacttttg tttttgatgt ctttcactat 420
actctgtctg gtaccagctt acagctttct gccgggagtc tactcagtct tttacttgta 480
cttgacattc tatttcacca atgatgtttc attcttggct caccttcaat ggtttgccat 540
gttttctcct attgtgcctt tttggataac agcaatctat gtattctgta tttctctgaa 600
gcactgccat tggttcttta acaactatct taggaaaaga gtcattgtta atggagttac 660
atttagtacc ttcgaggagg ctgctttgtg tacctttttg ctcaacaagg aaatgtacct 720
aaaattgcgt agcgagacac tgttgccact tacacagtat aacaggatc ttgctctata 780
taacaagtac aagtatttca gtggagcctt agatactacc agctatcgtg aagcagcttg 840
ctgccactta gcaaaggctc taaatgactt tagcaactca ggtgctgatg ttctctacca 900
accaccacag acatcaatca cttctgctgt tctgcagagt ggttttagga aaatggcatt 960
cccgtcaggc aaagttgaag ggtgcatggt acaagtaacc tgtggaacta caactcttaa 1020
tggattgtgg ttggatgaca cagtatactg tccaagacat gtcatttgca cagcagaaga 1080
catgcttaat cctaactatg aagatctgct cattcgcaa tccaaccata gctttcttgt 1140
tcaggctggc aatgttcaac ttcgtgttat tggccattct atgcaaaatt gtctgcttag 1200
gcttaaagtt gatacttcta accctaagac acccaagtat aaatttgtcc gtatccaacc 1260
tgggtcaaca ttttcagttc tagcatgcta caatggttca ccatctggtg tttatcagtg 1320
tgccatgaga cctaatacata ccattaaagg ttctttcctt aatggatcat gtggtagtgt 1380
tggttttaac attgattatg attgctgtc tttctgctat atgcatcata tggagcttcc 1440
aacaggagta cacgctggta ctgacttaga aggtaaattc tatggtccat ttgttgacag 1500
acaaactgca caggctgcag gtacagacac aaccataaca ttaaattgtt tggcatggct 1560
gtatgctgct gttatcaatg gtgatagggt gtttcttaat agattcacca ctactttgaa 1620
tgactttaac cttgtggcaa tgaagtacaa ctatgaacct ttgacacaag atcatgttga 1680
catattggga cctctttctg ctcaaacagg aattgccgtc ttagatatgt gtgctgcttt 1740
gaaagagctg ctgcagaatg gtatgaatgg tcgtactatc cttggtagca ctattttaga 1800
agatgagttt acaccatttg atgttgtag acaatgctct ggtgttacct tccaaggtaa 1860
gttcaagaaa attgttaagg gcactcatca ttggatgctt ttaactttct tgacatcact 1920
attgattctt gttcaaagta cacagtgggt actgtttttc tttgtttacg agaatgcttt 1980

```

S226CAS111.ST25

cttgccattt actcttggtt ttatggcaat tgctgcatgt	2020
<210> 49<211> 2040<212> DNA<213> CORONAVIRUS<400> 49	
agcattttcca gcctgaagac gtactgtagc agctaaactg cccagcacca tacctctatt	60
taggttggtt aagcctttga tgaagtacaa gtatttctact ttaggccctt ttggtgtgtc	120
tgtaacaaac ctacaagggtg gttccagttc tgtgtaaatt gtacctgtac catcactctt	180
agggaatcta gcccatttga gatcttggtg gtctgatagt aatgccagca caaacctacc	240
tcccttcgaa ttgttatagt aggcaagtgc attgtcatca gtacaagctg tttgtgtggt	300
accagccgca caggacatct gtcgtagtgc tactggactc agttcattat tctgtagttt	360
aacagctgag ttggctctta gagctgtaac aataagaggc caagccaaat ttggtgaatt	420
gtccatgtta atttctactaa gttgaacaat cttgctatcc gcatcaacaa cttgctggat	480
ttcccagagt gcagatgcat atgtaaaagg gttaccatca caagtgttct tgtaggtacc	540
ataatcaggg acaacaacca tgagtttggc tgctgtagtc aatggtatga tgttgagtgg	600
aacacaacca tcacgcgcat tgttgataat gttgttaagt gcatcattat caagcttcct	660
aagcatagtg aagagcattg tttgcatagc actagttact tttgccctct tgcctcaga	720
tcttgctgt ttgtacattt gggcatagc ctgatctgcc atcttttcca acttgcttg	780
catggcagca tcacggtcaa actcagattt agccacattc aaagatttct ttaacttttt	840
gagaacgact tcagaatcac cattagctac agcctgtcga taggcctcct gggcagtggc	900
ataagcggca tatgatggta aagaactaaa ttctgaagca atagcctgaa gagtagcacg	960
gttatcgagc atttcctcgc acaacctatt aatgtctaca gcaccctgca tggatagcaa	1020
aacagacaaa agagaaaacca tcttctcgaa agcttcagtt gtgtcttttg caagaagaat	1080
atcattgtgg agttgtacac attgtgcca caatttagaa gatgactcta ctctaagttg	1140
ttgaagaacc gagagcagta ccacagatgt gcactttacg tcagacattt tagactgtac	1200
agtagcaacc ttgatacatg gttacctcc aatacccaac aacttaatgt taagcttgaa	1260
agcatcaata ctactcttag gaggcaaaag cccctgggag ttcataatac taaattcttg	1320
tgtagagacc aagtagtcat aaacaccaag agtaagcctg aagtaacggt tgagtaaaca	1380
gaaaaggcca aagtagcagc agcaacaata gcctaagaaa caataaaca gcatgataca	1440
ctgtaagggtg ttgccagtaa taaataacaa tgggtaatac tcaacacaca caaacactat	1500
agctctagct aaaaacatga tagtcgtaac gacaccagaa tagttagagg ttacagaaat	1560
aactaaggcc cacatggaaa tagcttgatc taaagcatta ccatagtaga ctttgtaaac	1620
aagtgtaatg acattcatca gtgtccaaac acgtctagca gcatcatcat aaacagtgcg	1680
agctgtcatg agaataagca aaactaaagc tgaagcatac ataacacaat ccttaagcct	1740
ataaccagac aagctagtgt cagccaattc aagccatgtc atgatacgca tcaccagct	1800
agcaggcatg tagaccatat taaagtaagc aactgttgca agagaaggta acagaaacaa	1860
gcacaagaat gcgtgcttat gcttaacaag cagcatagca catgcagcaa ttgccataat	1920

S226CAS111.ST25

accaagagta aatggcaaga aagcattctc gtaaacaag aaaaacagtg accactgtgt	1980
actttgaaca agaatcaata gtgatgtcaa gaaagttaaa agcatccaat gatgagtgc	2040

<210> 50<211> 2012<212> DNA<213> CORONAVIRUS<400> 50	60
cttgtagggtt tgttacagac acaccaaag ggcctaaagt gaaatacttg tacttcatca	
aaggcttaaa caacctaaat agaggatatg tgctgggcag tttagctgct acagtacgtc	120
ttcaggctgg aaatgctaca gaagtacctg ccaattcaac tgtgctttcc ttctgtgctt	180
ttgcagtaga ccctgctaaa gcatataagg attacctagc aagtggagga caaccaatca	240
ccaactgtgt gaagatgttg tgtacacaca ctggtacagg acaggcaatt actgtaacac	300
cagaagctaa catggaccaa gagtcctttg gtggtgcttc atgttgtctg tattgtagat	360
gccacattga ccatccaaat cctaaaggat tctgtgactt gaaaggtaag tacgtccaaa	420
tacctaccac ttgtgctaata gaccagtggt gttttactact tagaaacaca gtctgtaccg	480
tctgcggaat gtggaaagggt tatggctgta gttgtgacca actccgcgaa cccttgatgc	540
agtctgcgga tgcataacg tttttaaacg gggttgcggt gtaagtgcag cccgtcttac	600
accgtgcggc acaggcacta gtactgatgt cgtctacagg gcttttgata tttacaacga	660
aaaagttgct gggtttgcaa agttcctaaa aactaattgc tgtcgcttcc aggagaagga	720
tgaggaaggc aattttattag actcttactt tgtagttaag aggcatacta tgtctaacta	780
ccaacatgaa gagactatatt ataacttgggt taaagattgt ccagcggttg ctgtccatga	840
ctttttcaag tttagagtag atggtgacat ggtaccacat atatcacgtc agcgtctaac	900
taaatacaca atggctgatt tagtctatgc tctacgtcat tttgatgagg gtaattgtga	960
tacattaaaa gaaatactcg tcacatacaa ttgctgtgat gatgattatt tcaataagaa	1020
ggattgggtat gacttcgtag agaatcctga catcttacgc gtatatgcta acttaggtga	1080
gcgtgtacgc caatcattat taaagactgt acaattctgc gatgctatgc gtgatgcagg	1140
cattgtaggc gtactgacat tagataatca ggatcttaat gggaactgggt acgatttcgg	1200
tgatttcgta caagtagcac caggctgcgg agttcctatt gtggattcat attactcatt	1260
gctgatgccc atcctcactt tgactagggc attggctgct gagtcccata tggatgctga	1320
tctcgcaaaa ccacttatta agtgggattt gctgaaatat gattttacgg aagagagact	1380
ttgtctcttc gaccgttatt ttaaattattg ggaccagaca taccatccca attgtattaa	1440
ctgtttggat gataggtgta tccttcattg tgcaaaacttt aatgtgttat tttctactgt	1500
gtttccacct acaagttttg gaccactagt aagaaaaata tttgtagatg gtgttccttt	1560
tgttgtttca actggatacc attttcgtga gttaggagtc gtacataatc aggatgtaaa	1620
cttacatagc tcgcgtctca gtttcaagga acttttagtg tatgctgctg atccagctat	1680
gcatgcagct tctggcaatt tattgctaga taaacgcact acatgctttt cagtagctgc	1740
actaacaac aatgttgctt ttcaaactgt caaaccgggt aattttaata aagactttta	1800
tgactttgct gtgtctaaag gtttctttaa ggaaggaagt tctgttgaac taaaacactt	1860



S226CAS111.ST25

cttcttttgct caggatggca acgctgctat cagtgattat gactattatc gttataatct	1920
gccaacaatg tgtgatatca gacaactcct attcgtagtt gaagttgttg ataaatactt	1980
tgattgttac gatggtggct gtattaatgc ca	2012
<210> 51<211> 1877<212> DNA<213> CORONAVIRUS<400> 51	
gtacttcgcg tacagtggca ataccatatg acagcttaaa tgtttcctca gtggctttga	60
gcgtttctgc tgcgaaaagc ttgagtctct cagtacaagt gttggcaagt atgtaatcgc	120
cagcattagt ccaatcacat gttgctatcg cattgaagtc agtgacattg tcaactgccta	180
cacatgtgtt tttgtataaa ccaaaaacct gaccattagc acataatgga aaactaatgg	240
gaggcttatg tgacttgcaa taatagctca tacctcctag atacagttgt gtcacatcag	300
tgacatcaca acctggggca ttgcaaacat agggattaac agacaacact aatttgtgtg	360
atgttgaaat gacatgggca tagcagcact tgcaacatag gaatgggtctc ctaatacagg	420
caccgcaacg aagtgaagtc tgtgaattgc acaatacaca agcacctaca gcctgcaaga	480
ctgtatgtgg tgtgtacata gcctcataaa actcagggttc ccagtaccgt gaggtgttat	540
cattagttag cattacggaa tacatgtcca acatgtggcc agtaagctca tcatgtaact	600
ttctaatagta ttgtaaatac aagtgaaga catcagcata ctcttgatta ggatgttttg	660
taagtgggta agcatcaata gccagtgaca cgaacctttc aatcataagt gtaccatctg	720
ttttgacaat atcatcgaca aaacagcctg cgcctaatat tcttgatgga tctgggtaag	780
gcagggtacac gtaatcatct ccttgtttaa ctagcattgt atgctgtgag caaaattcgt	840
gaggctcctt agtaagggtc gtctcagtc aacattttgc ctcagacatg aacacattat	900
tttgataata aagaactgcc ttaaagtctt taatgctagc tactaaacct tgagccgcat	960
agttactgtt atagcacaca acggcatcat cagaaagaat catcatggag aaatgtttac	1020
gcaggtaagc gtaaaactca tccacgaatt catgatcaac atccctatct ctatagagac	1080
actcatagag cctgtgttgt agattgcgga catacttgtc agctatctta ttaccatcag	1140
ttgaaagaag tgcatttaca ttggctgtaa cagcttgaca aatgttaaag acactattag	1200
cataagcagt tgtagcatca ccggatgatg ttccacctgg tttaacatat agtgagccgc	1260
cacacatgac catctcactt aatacttgcg cacactcgtt agctaacctg tagaaacggt	1320
gtgataagtt acagcaagtg ttatgtttgc gagcaagaac aagagaggcc attatcctaa	1380
gcatgttagg catggctctg tcacattttg gataatcca acccataagg tgtggagttt	1440
ctacatcact gtaaacagtt tttaacatat tatgccagcc accgtaaaac ttgcttgctc	1500
caattaccac agtagctcct ctagtggcgg ctattgactt caataatttc tgatgaaact	1560
gtctattttgt catagtacta cagatagaga caccagctac ggtgcgagct ctattccttg	1620
cactaatggc atacttaaga ttcatttgag ttatagtagg gatgacatta cgcttagtat	1680
acgcgaaaag tgcattctga tcctcataac tcattgagtc ataataaagt ctagccttac	1740
cccatttatt aaatgggaaa ccagctgatt tatccagatt gttaacgatt acttggttgg	1800

S226CAS111.ST25  
cattaataca gccaccatcg taacaatcaa agtattttatc aacaacttca actacgaata 1860  
ggagttgtct gatatca 1877

<210> 52<211> 2051<212> DNA<213> CORONAVIRUS<400> 52  
tcaggtccaa tcttgacaaa gtacttcatt gatgtaagct caaagccatg cgcccaaagg 60  
acgaacacga ctctgtctga caatcctttc agtgtatcac tgagcatttg tactatctta 120  
atagcacta cattccaggg caagccttta tacatgagtg gtataagatg tttaaactgg 180  
tcacctggtg gaggttttgc attaaactctg gtgaattctg tgttattttc agtgtcaaca 240  
taaccagtcg gtacagctac taagttaaca cctgtagaaa atcctagctg gagaggtagg 300  
ttagtaccga cagcatctct agttgcatga cagccctcta catcaaagcc aatccacgca 360  
cgaacgtgac gaatagcttc ttcgcgggtg ataaacatat tagggtaacc attgacttgg 420  
taattcattt tgaaacccat catagagatg agtctacggt aggtcatgtc ctttggtatg 480  
cctgggtatgt caacacataa tccttcagtc ttgaacttta tatcaacgct gaggtgtgta 540  
ggtgcctgtg taggatgaag accagtaatg atcttactac agtccttaaa aagtcagtt 600  
acattttctg cttgtaatgt agccacattg cgacgtggta tttctagact tgtaaattgc 660  
agtttgtcat aaagatctct atcagacatt atgcacaaaa tgccaatttt tgcccttgtg 720  
atagccacat tgaagcgggt gacattacaa gagtgtgctg tttcagtagt ttgtgtgaat 780  
atgacatagt catattcaga accctgtgat gaatcaacag tctgcgtagg caatcctaag 840  
atTTTTgaag ctacagcggt ctgtgaatta taagggtgaga taaaaacagc ttttctccaa 900  
gcaggattgc gtgtaagaaa ttctcttaca acgcctatTT gaggtctggt gattgcagat 960  
gaaacatcat gtgtaataac acctttgtag aacattttga agcattgagc tgacttatcc 1020  
ttgtgtgctt ttagcttatt gtcataaaact aaagcactca cagtgtcaac aatttcagca 1080  
ggacaacggc gacaagttcc aaggaacatg tctggaccta ttgttttcat aagtctgcac 1140  
actgaattaa aatattcttg ttctagtgtg ctttagtca gcaatgtgcg gggggctggt 1200  
aattgagcag gatcgccaat atagacgtag tgttttgcac gaagtctagc attgacaaca 1260  
ctcaagtcat aattagtagc catagagatt tcatcaaaga ctacaatgtc agcagttggt 1320  
tctggcaatg catttacagt gcagaaaaca tactgttcta gtgttgaatt cactttgaat 1380  
ttatcaaaac actctacgcg cgcacgcgca ggtatgattc tactacattt atctatgggc 1440  
aaatatttta atgccttttc acatagggca tcaacagctg catgagagca tgccgtatac 1500  
actatgcgag cagatgggta atagagagca agtccgatgg caaaatgaCT cttaccagta 1560  
ccaggtggtc cttggagtgt agagtacttt tgcattgccg ctttttgata atttgcaaca 1620  
ttgctagaaa actcatctga gatgttgagt gttgggtaca agccagtaat tctcacatag 1680  
tgctcttgtg gcactagagt aggtgcacta agtggcatta cagtgtgaga tgtcaacaca 1740  
aagtaatcac caacattcaa cttgtatgtc gtagtacctc tgtacacaac agcatcacca 1800  
tagtcacctt tttcaaagggt gtactctcca atctgtactt tactattttt agttacacgg 1860

S226CAS111.ST25

taaccagtaa agacatagtt tctgttcaat ggtggtctag gttttccaac ctcccatgaa	1920
agatgcaatt ctctgtcaga gagtacttcg cgtacagtgg caataccata tgacagctta	1980
aatgtttcct cagtggcttt gagcgtttct gctgcgaaaa gcttgagtct ctcagtacaa	2040
gtgttgca g	2051

<210> 53<211> 2075<212> DNA<213> CORONAVIRUS<400> 53	
tgctttagt tttgggtaga aggtttcaac atgtccatcc ttacaccaa gcatgaatga	60
aatttcagca tagtcaattg taaccttgac cacttttgaa atcactgaca aatcttgtga	120
ctttattatc tcgacaaagt catcaagtaa aagatcaatc acagaacaca cacattttga	180
tgaacctgtt tgcgcatctg ttatgaagta atttttcact gtgctgtcca tagggataaa	240
atcctcta at ttaagtgggtg aatcttgtga gcgcttggt aagcctatca ttaaataag	300
accgccaagt tgtccatgac tgaaatctcc ataaacgatg tgttcgaagg catagccctc	360
gagcttatat cgctgtatga attcatccat agcgagctcg agaaagtcag tttccatttg	420
tgatctgggc ttaaaatcct ctaagtctct gctctgagta aagtaggttt caggcaactg	480
ttgaataatg ccgtctactt tcttaaagta gttaaactgt gtttttactg attctccaat	540
taatgtgact ccattgacgc tagcttgtgc tgggtccctt gaagggtgta gaccttgac	600
tgaaccttct gttattaaaa caccattacg ggcgtttcta aaaagggtcta cctgtccttc	660
cactctacca tcaaacaaga cagtaagtga agaacaagca ctctcagtag gtttcttggc	720
aatgtcagtc attgtgcaga cacctattgt agatacatgt gctggggctt ctcttttgta	780
gtcccagatt acagtattag cagcgatatc aacacccaaa ttattgagta tcttaatctc	840
tggcactggt ttaatgttac gcttagccca aagctcaaat gcaacattaa caggaagtgt	900
tgtcttattt tcaaagatct ccacatcaat accatctacc tttgtgtaaa cagcattatt	960
aatgatggaa acaggtgctt cgccggcggtg tccatcaaag tgtcctttat taacaacatt	1020
ataagccaca ttttctaacc tctgtaacct ggtaaatgta ttccacaggt tataagtatc	1080
aaattgtttg taaatccata ggctaaatcc agcagaaatc atcatattat atgcatccaa	1140
gtactgtcgg tactcatttg catggtgtct gcaaacagca ccacctaaat tgcatcgtgt	1200
aatacacgta gcagatttga gtggaacata atcaatatcc gacactactt gtttgccatg	1260
agactcaciaa ggactatcag aatagtaaaa gaaaggcaat tgctttaaat tagtaaatgc	1320
acttttatcg aaagctggag tgtggaatgc atgcttattc acatacaaac taccaccatc	1380
acagcctggt aagttcaagt ttgacaagac tcttgtgtca aacctacaca caattgcatt	1440
ggctgggtaa cgatcaacgt tacaattcca aaacaaacaa acaccatcag tgaatttatc	1500
gtgatgtgta gcataagaat agaagagttc ctctattttg taagctttgt cactacatgg	1560
ctgagcatcg tagaacttcc attctacttc agcctgaggc acacacttga tagcctttgg	1620
atttccaatg tcatgaagaa ctggaaactt atcagcaagc aatgcagact tcacaaccat	1680
gtgttgtact tttctgcaag cagaattaac cctcagttca tctcctataa taggggtattc	1740

S226CAS111.ST25

aacagaccaa tcaacgcgct taacaaagca ctcattggact gctaaacatc tagtcatgat	1800
agcatcacaa ctaggccacat gtgcattttcc atgtacctgg caatgttggt catgggttact	1860
ctgaagggtta cccgtaaagc cccactgctg aacatcaatc ataaatgggt tatagacata	1920
gtcaaaaccc acagaatgat tccagcaggc ataagtatct gatgaagtag aaaagcaagt	1980
tgcacgtttg tcacacagac aacacgttct ttcagggtcca atcttgacaa agtacttcat	2040
tgatgtaagc tcaaagccat gcgcccgaag gacga	2075

<210> 54<211> 1891<212> DNA<213> CORONAVIRUS<400> 54	
aagattcacc acttaaatta gaggatttta tccctatgga cagcacagtg aaaaattact	60
tcataacaga tgcgcaaaca gggtcatcaa aatgtgtgtg ttctgtgatt gatcttttac	120
ttgatgactt tgtcgagata ataaagtcac aagatttgtc agtgatttca aaagtgggtca	180
aggttacaat tgactatgct gaaatttcat tcatgctttg gtgtaaggat ggacatgttg	240
aaaccttcta cccaaaacta caagcaagtc aagcgtggca accagggtgtt gcgatgccta	300
acttgtaaa gatgcaaaga atgcttcttg aaaagtgtga ccttcagaat tatggtgaaa	360
atgctgttat accaaaagga ataattgatga atgtcgcaaa gtatactcaa ctgtgtcaat	420
acttaaatc acttacttta gctgtaccct acaacatgag agttattcac tttgggtgctg	480
gctctgataa aggagttgca ccagggtacag ctgtgtctcag acaatgggtg ccaactggca	540
cactacttgt cgattcagat cttaatgact tcgtctccga cgcagattct actttaattg	600
gagactgtgc aacagtacat acggctaata aatgggacct tattattagc gatatgtatg	660
accctaggac caaacatgtg acaaaagaga atgactctaa agaaggggtt ttcacttatt	720
tgtgtggatt tataaagcaa aaactagccc tgggtgggtc tatagctgta aagataacag	780
agcattcttg gaatgctgac ctttacaagc ttatgggcca tttctcatgg tggacagctt	840
ttgtttacaa tgtaaatgca tcatcatcgg aagcattttt aattggggct aactatcttg	900
gcaagccgaa ggaacaaatt gatggctata ccatgcatgc taactacatt ttctggagga	960
acacaaatcc tatccagttg tcttcctatt cactctttga catgagcaaa tttcctctta	1020
aattaagagg aactgctgta atgtctctta aggagaatca aatcaatgat atgatttatt	1080
ctcttctgga aaaaggtagg cttatcatta gagaaaacaa cagagttgtg gtttcaagtg	1140
atattcttgt taacaactaa acgaacatgt ttattttctt attatttctt actctcacta	1200
gtggtagtga ccttgaccgg tgcaccactt ttgatgatgt tcaagctcct aattacactc	1260
aacatacttc atctatgagg ggggtttact atcctgatga aatttttaga tcagacactc	1320
tttatttaac tcaggattta tttcttccat tttattctaa tgttacaggg tttcatacta	1380
ttaatcatac gtttggaac cctgtcatac cttttaagga tggattttat tttgctgcca	1440
cagagaaatc aaatgttgtc cgtgggtggg tttttgggtc taccatgaac aacaagtcac	1500
agtcggtgat tattattaac aattctacta atgttggtat acgagcatgt aactttgaat	1560
tgtgtgacaa ccctttcttt gctgtttcta aaccatggg tacacagaca catactatga	1620



S226CAS111.ST25

tattcgataa tgcatttaaat tgcacttttcg agtacatatc tgatgccttt tcgcttgatg	1680
tttcagaaaa gtcaggtaaat tttaaactac tacgagagtt tgtgtttaaa aataaagatg	1740
ggttttctcta tgtttataag ggctatcaac ctatagatgt agttcgtgat ctaccttctg	1800
gttttaacac ttgaaacct atttttaagt tgcctcttgg tattaacatt acaaatttta	1860
gagccattct tacagccttt tcacctgctc a	1891
<210> 55<211> 32<212> DNA<213> ARTIFICIAL SEQUENCE<220><223> amorce N sens	
<400> 55 cccatatgtc tgataatgga cccaatcaa ac	32
<210> 56<211> 32<212> DNA<213> artificial sequence<220><223> amorce N antisens	
<400> 56 ccccgggtg cctgagttga atcagcagaa gc	32
<210> 57<211> 31<212> DNA<213> artificial sequence<220><223> amorce Sc sens	
<400> 57 cccatatgag tgaccttgac cggtgcacca c	31
<210> 58<211> 30<212> DNA<213> artificial sequence<220><223> amorce SL sens	
<400> 58 cccatatgaa accttgacc ccacctgctc	30
<210> 59<211> 33<212> DNA<213> amorce Sc et SL antisens<400> 59	33
ccccggggtt taatatattg ctcatatattt ccc	
<210> 60<211> 16<212> DNA<213> amorce sens série 1<400> 60	16
ggcatcgtat gggttg	
<210> 61<211> 16<212> DNA<213> amorce antisens série 2 (28774-28759)<400> 61	16
cagtttcacc acctcc	
<210> 62<211> 16<212> DNA<213> amorce sens série 2 (28375-28390)<400> 62	16
ggctactacc gaagag	
<210> 63<211> 16<212> DNA<213> amorce antisens série 2 (28702-28687)<400> 63	16
aattaccgcg actacg	
<210> 64<211> 26<212> DNA<213> sonde 1/série 1 (28561-28586)<400> 64	26
ggcaccgcga atcctaataa caatgc	
<210> 65<211> 21<212> DNA<213> sonde 2/série 1 (28588-28608)<400> 65	21
gccaccgtgc tacaacttcc t	
<210> 66<211> 23<212> DNA<213> sonde 1/série 2 /sonde N/FL (28541-28563)<400> 66	

atacacccaa agaccacatt ggc

23

<210> 67<211> 25<212> DNA<213> sonde 2/série 2/sonde SARS/N/LC705  
 (28565-28589)<400> 67  
 cccgcaatcc taataacaat gctgc

25

<210> 68<211> 30<212> DNA<213> artificial sequence<220><223> amorce ancre  
 14T

<400> 68  
 agatgaattc ggtacctttt tttttttttt

30

<210> 69<211> 13<212> PRT<213> artificial sequence<220><223> peptide M2-14  
 <400> 69

Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln  
 1 5 10

<210> 70<211> 12<212> PRT<213> artificial sequence<220><223> peptide E1-12  
 <400> 70

Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu  
 1 5 10

<210> 71<211> 24<212> PRT<213> artificial sequence<220><223> peptide E53-72  
 <400> 71

Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn Leu Asn Ser Ser  
 1 5 10 15

Glu Gly Val Pro Asp Leu Leu Val  
 20

<210> 72<211> 153<212> DNA<213> CORONAVIRUS<400> 72  
 gatattaggt ttttacctac ccaggaaaag ccaaccaacc tcgatctctt gtagatctgt 60  
 tctctaaacg aacttttaaaa tctgtgtagc tgtcgctcgg ctgcatgcct agtgcaccta 120  
 cgcagtataa acaataataa attttactgt cgt 153

<210> 73<211> 410<212> DNA<213> CORONAVIRUS<400> 73  
 ttctccagac aacttcaaaa ttccatgagt ggagcttctg ctgattcaac tcaggcataa 60  
 acactcatga tgaccacaca aggcatgagg gctatgtaaa cgttttcgca attccgttta 120  
 cgatacatag tctactcttg tgcagaatga attctcgtaa ctaaacagca caagtaggtt 180  
 tagttaactt taatctcaca tagcaatctt taatcaatgt gtaacattag ggaggacttg 240  
 aaagagccac cacattttca tcgaggccac gcggagtacg atcgagggtg cagtgaataa 300  
 tgctagggag agctgcctat atggaagagc cctaattgtgt aaaattaatt ttagtagtgc 360  
 tatccccatg tgatttttaat agcttcttag gagaatgaca aaaaaaaaaa 410

<210> 74<211> 4382<212> PRT<213> CORONAVIRUS<400> 74  
 Page 75

S226CAS111.ST25

Met Glu Ser Leu Val Leu Gly Val Asn Glu Lys Thr His Val Gln Leu  
1 5 10 15

Ser Leu Pro Val Leu Gln Val Arg Asp Val Leu Val Arg Gly Phe Gly  
20 25 30

Asp Ser Val Glu Glu Ala Leu Ser Glu Ala Arg Glu His Leu Lys Asn  
35 40 45

Gly Thr Cys Gly Leu Val Glu Leu Glu Lys Gly Val Leu Pro Gln Leu  
50 55 60

Glu Gln Pro Tyr Val Phe Ile Lys Arg Ser Asp Ala Leu Ser Thr Asn  
65 70 75 80

His Gly His Lys Val Val Glu Leu Val Ala Glu Met Asp Gly Ile Gln  
85 90 95

Tyr Gly Arg Ser Gly Ile Thr Leu Gly Val Leu Val Pro His Val Gly  
100 105 110

Glu Thr Pro Ile Ala Tyr Arg Asn Val Leu Leu Arg Lys Asn Gly Asn  
115 120 125

Lys Gly Ala Gly Gly His Ser Tyr Gly Ile Asp Leu Lys Ser Tyr Asp  
130 135 140

Leu Gly Asp Glu Leu Gly Thr Asp Pro Ile Glu Asp Tyr Glu Gln Asn  
145 150 155 160

Trp Asn Thr Lys His Gly Ser Gly Ala Leu Arg Glu Leu Thr Arg Glu  
165 170 175

Leu Asn Gly Gly Ala Val Thr Arg Tyr Val Asp Asn Asn Phe Cys Gly  
180 185 190

Pro Asp Gly Tyr Pro Leu Asp Cys Ile Lys Asp Phe Leu Ala Arg Ala  
195 200 205

Gly Lys Ser Met Cys Thr Leu Ser Glu Gln Leu Asp Tyr Ile Glu Ser  
210 215 220

Lys Arg Gly Val Tyr Cys Cys Arg Asp His Glu His Glu Ile Ala Trp  
225 230 235 240

Phe Thr Glu Arg Ser Asp Lys Ser Tyr Glu His Gln Thr Pro Phe Glu  
245 250 255

Ile Lys Ser Ala Lys Lys Phe Asp Thr Phe Lys Gly Glu Cys Pro Lys  
260 265 270

S226CAS111.ST25

Phe Val Phe Pro Leu Asn Ser Lys Val Lys Val Ile Gln Pro Arg Val  
 275 280 285  
 Glu Lys Lys Lys Thr Glu Gly Phe Met Gly Arg Ile Arg Ser Val Tyr  
 290 295 300  
 Pro Val Ala Ser Pro Gln Glu Cys Asn Asn Met His Leu Ser Thr Leu  
 305 310 315 320  
 Met Lys Cys Asn His Cys Asp Glu Val Ser Trp Gln Thr Cys Asp Phe  
 325 330 335  
 Leu Lys Ala Thr Cys Glu His Cys Gly Thr Glu Asn Leu Val Ile Glu  
 340 345 350  
 Gly Pro Thr Thr Cys Gly Tyr Leu Pro Thr Asn Ala Val Val Lys Met  
 355 360 365  
 Pro Cys Pro Ala Cys Gln Asp Pro Glu Ile Gly Pro Glu His Ser Val  
 370 375 380  
 Ala Asp Tyr His Asn His Ser Asn Ile Glu Thr Arg Leu Arg Lys Gly  
 385 390 395 400  
 Gly Arg Thr Arg Cys Phe Gly Gly Cys Val Phe Ala Tyr Val Gly Cys  
 405 410 415  
 Tyr Asn Lys Arg Ala Tyr Trp Val Pro Arg Ala Ser Ala Asp Ile Gly  
 420 425 430  
 Ser Gly His Thr Gly Ile Thr Gly Asp Asn Val Glu Thr Leu Asn Glu  
 435 440 445  
 Asp Leu Leu Glu Ile Leu Ser Arg Glu Arg Val Asn Ile Asn Ile Val  
 450 455 460  
 Gly Asp Phe His Leu Asn Glu Glu Val Ala Ile Ile Leu Ala Ser Phe  
 465 470 475 480  
 Ser Ala Ser Thr Ser Ala Phe Ile Asp Thr Ile Lys Ser Leu Asp Tyr  
 485 490 495  
 Lys Ser Phe Lys Thr Ile Val Glu Ser Cys Gly Asn Tyr Lys Val Thr  
 500 505 510  
 Lys Gly Lys Pro Val Lys Gly Ala Trp Asn Ile Gly Gln Gln Arg Ser  
 515 520 525  
 Val Leu Thr Pro Leu Cys Gly Phe Pro Ser Gln Ala Ala Gly Val Ile  
 530 535 540

S226CAS111.ST25

Arg Ser Ile Phe Ala Arg Thr Leu Asp Ala Ala Asn His Ser Ile Pro  
545 550 555 560

Asp Leu Gln Arg Ala Ala Val Thr Ile Leu Asp Gly Ile Ser Glu Gln  
565 570 575

Ser Leu Arg Leu Val Asp Ala Met Val Tyr Thr Ser Asp Leu Leu Thr  
580 585 590

Asn Ser Val Ile Ile Met Ala Tyr Val Thr Gly Gly Leu Val Gln Gln  
595 600 605

Thr Ser Gln Trp Leu Ser Asn Leu Leu Gly Thr Thr Val Glu Lys Leu  
610 615 620

Arg Pro Ile Phe Glu Trp Ile Glu Ala Lys Leu Ser Ala Gly Val Glu  
625 630 635 640

Phe Leu Lys Asp Ala Trp Glu Ile Leu Lys Phe Leu Ile Thr Gly Val  
645 650 655

Phe Asp Ile Val Lys Gly Gln Ile Gln Val Ala Ser Asp Asn Ile Lys  
660 665 670

Asp Cys Val Lys Cys Phe Ile Asp Val Val Asn Lys Ala Leu Glu Met  
675 680 685

Cys Ile Asp Gln Val Thr Ile Ala Gly Ala Lys Leu Arg Ser Leu Asn  
690 695 700

Leu Gly Glu Val Phe Ile Ala Gln Ser Lys Gly Leu Tyr Arg Gln Cys  
705 710 715 720

Ile Arg Gly Lys Glu Gln Leu Gln Leu Leu Met Pro Leu Lys Ala Pro  
725 730 735

Lys Glu Val Thr Phe Leu Glu Gly Asp Ser His Asp Thr Val Leu Thr  
740 745 750

Ser Glu Glu Val Val Leu Lys Asn Gly Glu Leu Glu Ala Leu Glu Thr  
755 760 765

Pro Val Asp Ser Phe Thr Asn Gly Ala Ile Val Gly Thr Pro Val Cys  
770 775 780

Val Asn Gly Leu Met Leu Leu Glu Ile Lys Asp Lys Glu Gln Tyr Cys  
785 790 795 800

Ala Leu Ser Pro Gly Leu Leu Ala Thr Asn Asn Val Phe Arg Leu Lys  
805 810 815

## S226CAS111.ST25

Gly Gly Ala Pro Ile Lys Gly Val Thr Phe Gly Glu Asp Thr Val Trp  
 820 825 830  
 Glu Val Gln Gly Tyr Lys Asn Val Arg Ile Thr Phe Glu Leu Asp Glu  
 835 840 845  
 Arg Val Asp Lys Val Leu Asn Glu Lys Cys Ser Val Tyr Thr Val Glu  
 850 855 860  
 Ser Gly Thr Glu Val Thr Glu Phe Ala Cys Val Val Ala Glu Ala Val  
 865 870 875 880  
 Val Lys Thr Leu Gln Pro Val Ser Asp Leu Leu Thr Asn Met Gly Ile  
 885 890 895  
 Asp Leu Asp Glu Trp Ser Val Ala Thr Phe Tyr Leu Phe Asp Asp Ala  
 900 905 910  
 Gly Glu Glu Asn Phe Ser Ser Arg Met Tyr Cys Ser Phe Tyr Pro Pro  
 915 920 925  
 Asp Glu Glu Glu Glu Asp Asp Ala Glu Cys Glu Glu Glu Ile Asp  
 930 935 940  
 Glu Thr Cys Glu His Glu Tyr Gly Thr Glu Asp Asp Tyr Gln Gly Leu  
 945 950 955 960  
 Pro Leu Glu Phe Gly Ala Ser Ala Glu Thr Val Arg Val Glu Glu Glu  
 965 970 975  
 Glu Glu Glu Asp Trp Leu Asp Asp Thr Thr Glu Gln Ser Glu Ile Glu  
 980 985 990  
 Pro Glu Pro Glu Pro Thr Pro Glu Glu Pro Val Asn Gln Phe Thr Gly  
 995 1000 1005  
 Tyr Leu Lys Leu Thr Asp Asn Val Ala Ile Lys Cys Val Asp Ile  
 1010 1015 1020  
 Val Lys Glu Ala Gln Ser Ala Asn Pro Met Val Ile Val Asn Ala  
 1025 1030 1035  
 Ala Asn Ile His Leu Lys His Gly Gly Gly Val Ala Gly Ala Leu  
 1040 1045 1050  
 Asn Lys Ala Thr Asn Gly Ala Met Gln Lys Glu Ser Asp Asp Tyr  
 1055 1060 1065  
 Ile Lys Leu Asn Gly Pro Leu Thr Val Gly Gly Ser Cys Leu Leu  
 1070 1075 1080

## S226CAS111.ST25

Ser Gly His Asn Leu Ala Lys Lys Cys Leu His Val Val Gly Pro  
 1085 1090 1095  
 Asn Leu Asn Ala Gly Glu Asp Ile Gln Leu Leu Lys Ala Ala Tyr  
 1100 1105 1110  
 Glu Asn Phe Asn Ser Gln Asp Ile Leu Leu Ala Pro Leu Leu Ser  
 1115 1120 1125  
 Ala Gly Ile Phe Gly Ala Lys Pro Leu Gln Ser Leu Gln Val Cys  
 1130 1135 1140  
 Val Gln Thr Val Arg Thr Gln Val Tyr Ile Ala Val Asn Asp Lys  
 1145 1150 1155  
 Ala Leu Tyr Glu Gln Val Val Met Asp Tyr Leu Asp Asn Leu Lys  
 1160 1165 1170  
 Pro Arg Val Glu Ala Pro Lys Gln Glu Glu Pro Pro Asn Thr Glu  
 1175 1180 1185  
 Asp Ser Lys Thr Glu Glu Lys Ser Val Val Gln Lys Pro Val Asp  
 1190 1195 1200  
 Val Lys Pro Lys Ile Lys Ala Cys Ile Asp Glu Val Thr Thr Thr  
 1205 1210 1215  
 Leu Glu Glu Thr Lys Phe Leu Thr Asn Lys Leu Leu Leu Phe Ala  
 1220 1225 1230  
 Asp Ile Asn Gly Lys Leu Tyr His Asp Ser Gln Asn Met Leu Arg  
 1235 1240 1245  
 Gly Glu Asp Met Ser Phe Leu Glu Lys Asp Ala Pro Tyr Met Val  
 1250 1255 1260  
 Gly Asp Val Ile Thr Ser Gly Asp Ile Thr Cys Val Val Ile Pro  
 1265 1270 1275  
 Ser Lys Lys Ala Gly Gly Thr Thr Glu Met Leu Ser Arg Ala Leu  
 1280 1285 1290  
 Lys Lys Val Pro Val Asp Glu Tyr Ile Thr Thr Tyr Pro Gly Gln  
 1295 1300 1305  
 Gly Cys Ala Gly Tyr Thr Leu Glu Glu Ala Lys Thr Ala Leu Lys  
 1310 1315 1320  
 Lys Cys Lys Ser Ala Phe Tyr Val Leu Pro Ser Glu Ala Pro Asn  
 1325 1330 1335

## S226CAS111.ST25

Ala Lys Glu Glu Ile Leu Gly Thr Val Ser Trp Asn Leu Arg Glu  
 1340 1345 1350  
 Met Leu Ala His Ala Glu Glu Thr Arg Lys Leu Met Pro Ile Cys  
 1355 1360 1365  
 Met Asp Val Arg Ala Ile Met Ala Thr Ile Gln Arg Lys Tyr Lys  
 1370 1375 1380  
 Gly Ile Lys Ile Gln Glu Gly Ile Val Asp Tyr Gly Val Arg Phe  
 1385 1390 1395  
 Phe Phe Tyr Thr Ser Lys Glu Pro Val Ala Ser Ile Ile Thr Lys  
 1400 1405 1410  
 Leu Asn Ser Leu Asn Glu Pro Leu Val Thr Met Pro Ile Gly Tyr  
 1415 1420 1425  
 Val Thr His Gly Phe Asn Leu Glu Glu Ala Ala Arg Cys Met Arg  
 1430 1435 1440  
 Ser Leu Lys Ala Pro Ala Val Val Ser Val Ser Ser Pro Asp Ala  
 1445 1450 1455  
 Val Thr Thr Tyr Asn Gly Tyr Leu Thr Ser Ser Ser Lys Thr Ser  
 1460 1465 1470  
 Glu Glu His Phe Val Glu Thr Val Ser Leu Ala Gly Ser Tyr Arg  
 1475 1480 1485  
 Asp Trp Ser Tyr Ser Gly Gln Arg Thr Glu Leu Gly Val Glu Phe  
 1490 1495 1500  
 Leu Lys Arg Gly Asp Lys Ile Val Tyr His Thr Leu Glu Ser Pro  
 1505 1510 1515  
 Val Glu Phe His Leu Asp Gly Glu Val Leu Ser Leu Asp Lys Leu  
 1520 1525 1530  
 Lys Ser Leu Leu Ser Leu Arg Glu Val Lys Thr Ile Lys Val Phe  
 1535 1540 1545  
 Thr Thr Val Asp Asn Thr Asn Leu His Thr Gln Leu Val Asp Met  
 1550 1555 1560  
 Ser Met Thr Tyr Gly Gln Gln Phe Gly Pro Thr Tyr Leu Asp Gly  
 1565 1570 1575  
 Ala Asp Val Thr Lys Ile Lys Pro His Val Asn His Glu Gly Lys  
 1580 1585 1590



## S226CAS111.ST25

Thr Phe Phe Val Leu Pro Ser Asp Asp Thr Leu Arg Ser Glu Ala  
 1595 1600 1605  
 Phe Glu Tyr Tyr His Thr Leu Asp Glu Ser Phe Leu Gly Arg Tyr  
 1610 1615 1620  
 Met Ser Ala Leu Asn His Thr Lys Lys Trp Lys Phe Pro Gln Val  
 1625 1630 1635  
 Gly Gly Leu Thr Ser Ile Lys Trp Ala Asp Asn Asn Cys Tyr Leu  
 1640 1650  
 Ser Ser Val Leu Leu Ala Leu Gln Gln Leu Glu Val Lys Phe Asn  
 1655 1660 1665  
 Ala Pro Ala Leu Gln Glu Ala Tyr Tyr Arg Ala Arg Ala Gly Asp  
 1670 1675 1680  
 Ala Ala Asn Phe Cys Ala Leu Ile Leu Ala Tyr Ser Asn Lys Thr  
 1685 1690 1695  
 Val Gly Glu Leu Gly Asp Val Arg Glu Thr Met Thr His Leu Leu  
 1700 1705 1710  
 Gln His Ala Asn Leu Glu Ser Ala Lys Arg Val Leu Asn Val Val  
 1715 1720 1725  
 Cys Lys His Cys Gly Gln Lys Thr Thr Thr Leu Thr Gly Val Glu  
 1730 1735 1740  
 Ala Val Met Tyr Met Gly Thr Leu Ser Tyr Asp Asn Leu Lys Thr  
 1745 1750 1755  
 Gly Val Ser Ile Pro Cys Val Cys Gly Arg Asp Ala Thr Gln Tyr  
 1760 1765 1770  
 Leu Val Gln Gln Glu Ser Ser Phe Val Met Met Ser Ala Pro Pro  
 1775 1780 1785  
 Ala Glu Tyr Lys Leu Gln Gln Gly Thr Phe Leu Cys Ala Asn Glu  
 1790 1795 1800  
 Tyr Thr Gly Asn Tyr Gln Cys Gly His Tyr Thr His Ile Thr Ala  
 1805 1810 1815  
 Lys Glu Thr Leu Tyr Arg Ile Asp Gly Ala His Leu Thr Lys Met  
 1820 1825 1830  
 Ser Glu Tyr Lys Gly Pro Val Thr Asp Val Phe Tyr Lys Glu Thr  
 1835 1840 1845

## S226CAS111.ST25

Ser Tyr Thr Thr Thr Ile Lys Pro Val Ser Tyr Lys Leu Asp Gly  
 1850 1855 1860  
 Val Thr Tyr Thr Glu Ile Glu Pro Lys Leu Asp Gly Tyr Tyr Lys  
 1865 1870 1875  
 Lys Asp Asn Ala Tyr Tyr Thr Glu Gln Pro Ile Asp Leu Val Pro  
 1880 1885 1890  
 Thr Gln Pro Leu Pro Asn Ala Ser Phe Asp Asn Phe Lys Leu Thr  
 1895 1900 1905  
 Cys Ser Asn Thr Lys Phe Ala Asp Asp Leu Asn Gln Met Thr Gly  
 1910 1915 1920  
 Phe Thr Lys Pro Ala Ser Arg Glu Leu Ser Val Thr Phe Phe Pro  
 1925 1930 1935  
 Asp Leu Asn Gly Asp Val Val Ala Ile Asp Tyr Arg His Tyr Ser  
 1940 1945 1950  
 Ala Ser Phe Lys Lys Gly Ala Lys Leu Leu His Lys Pro Ile Val  
 1955 1960 1965  
 Trp His Ile Asn Gln Ala Thr Thr Lys Thr Thr Phe Lys Pro Asn  
 1970 1975 1980  
 Thr Trp Cys Leu Arg Cys Leu Trp Ser Thr Lys Pro Val Asp Thr  
 1985 1990 1995  
 Ser Asn Ser Phe Glu Val Leu Ala Val Glu Asp Thr Gln Gly Met  
 2000 2005 2010  
 Asp Asn Leu Ala Cys Glu Ser Gln Gln Pro Thr Ser Glu Glu Val  
 2015 2020 2025  
 Val Glu Asn Pro Thr Ile Gln Lys Glu Val Ile Glu Cys Asp Val  
 2030 2035 2040  
 Lys Thr Thr Glu Val Val Gly Asn Val Ile Leu Lys Pro Ser Asp  
 2045 2050 2055  
 Glu Gly Val Lys Val Thr Gln Glu Leu Gly His Glu Asp Leu Met  
 2060 2065 2070  
 Ala Ala Tyr Val Glu Asn Thr Ser Ile Thr Ile Lys Lys Pro Asn  
 2075 2080 2085  
 Glu Leu Ser Leu Ala Leu Gly Leu Lys Thr Ile Ala Thr His Gly  
 2090 2095 2100

S226CAS111.ST25

Ile Ala Ala Ile Asn Ser Val Pro Trp Ser Lys Ile Leu Ala Tyr  
2105 2110 2115

Val Lys Pro Phe Leu Gly Gln Ala Ala Ile Thr Thr Ser Asn Cys  
2120 2125 2130

Ala Lys Arg Leu Ala Gln Arg Val Phe Asn Asn Tyr Met Pro Tyr  
2135 2140 2145

Val Phe Thr Leu Leu Phe Gln Leu Cys Thr Phe Thr Lys Ser Thr  
2150 2155 2160

Asn Ser Arg Ile Arg Ala Ser Leu Pro Thr Thr Ile Ala Lys Asn  
2165 2170 2175

Ser Val Lys Ser Val Ala Lys Leu Cys Leu Asp Ala Gly Ile Asn  
2180 2185 2190

Tyr Val Lys Ser Pro Lys Phe Ser Lys Leu Phe Thr Ile Ala Met  
2195 2200 2205

Trp Leu Leu Leu Leu Ser Ile Cys Leu Gly Ser Leu Ile Cys Val  
2210 2215 2220

Thr Ala Ala Phe Gly Val Leu Leu Ser Asn Phe Gly Ala Pro Ser  
2225 2230 2235

Tyr Cys Asn Gly Val Arg Glu Leu Tyr Leu Asn Ser Ser Asn Val  
2240 2245 2250

Thr Thr Met Asp Phe Cys Glu Gly Ser Phe Pro Cys Ser Ile Cys  
2255 2260 2265

Leu Ser Gly Leu Asp Ser Leu Asp Ser Tyr Pro Ala Leu Glu Thr  
2270 2275 2280

Ile Gln Val Thr Ile Ser Ser Tyr Lys Leu Asp Leu Thr Ile Leu  
2285 2290 2295

Gly Leu Ala Ala Glu Trp Val Leu Ala Tyr Met Leu Phe Thr Lys  
2300 2305 2310

Phe Phe Tyr Leu Leu Gly Leu Ser Ala Ile Met Gln Val Phe Phe  
2315 2320 2325

Gly Tyr Phe Ala Ser His Phe Ile Ser Asn Ser Trp Leu Met Trp  
2330 2335 2340

Phe Ile Ile Ser Ile Val Gln Met Ala Pro Val Ser Ala Met Val  
2345 2350 2355

## S226CAS111.ST25

Arg Met Tyr Ile Phe Phe Ala Ser Phe Tyr Tyr Ile Trp Lys Ser  
 2360 2365 2370  
 Tyr Val His Ile Met Asp Gly Cys Thr Ser Ser Thr Cys Met Met  
 2375 2380 2385  
 Cys Tyr Lys Arg Asn Arg Ala Thr Arg Val Glu Cys Thr Thr Ile  
 2390 2395 2400  
 Val Asn Gly Met Lys Arg Ser Phe Tyr Val Tyr Ala Asn Gly Gly  
 2405 2410 2415  
 Arg Gly Phe Cys Lys Thr His Asn Trp Asn Cys Leu Asn Cys Asp  
 2420 2425 2430  
 Thr Phe Cys Thr Gly Ser Thr Phe Ile Ser Asp Glu Val Ala Arg  
 2435 2440 2445  
 Asp Leu Ser Leu Gln Phe Lys Arg Pro Ile Asn Pro Thr Asp Gln  
 2450 2455 2460  
 Ser Ser Tyr Ile Val Asp Ser Val Ala Val Lys Asn Gly Ala Leu  
 2465 2470 2475  
 His Leu Tyr Phe Asp Lys Ala Gly Gln Lys Thr Tyr Glu Arg His  
 2480 2485 2490  
 Pro Leu Ser His Phe Val Asn Leu Asp Asn Leu Arg Ala Asn Asn  
 2495 2500 2505  
 Thr Lys Gly Ser Leu Pro Ile Asn Val Ile Val Phe Asp Gly Lys  
 2510 2515 2520  
 Ser Lys Cys Asp Glu Ser Ala Ser Lys Ser Ala Ser Val Tyr Tyr  
 2525 2530 2535  
 Ser Gln Leu Met Cys Gln Pro Ile Leu Leu Leu Asp Gln Ala Leu  
 2540 2545 2550  
 Val Ser Asp Val Gly Asp Ser Thr Glu Val Ser Val Lys Met Phe  
 2555 2560 2565  
 Asp Ala Tyr Val Asp Thr Phe Ser Ala Thr Phe Ser Val Pro Met  
 2570 2575 2580  
 Glu Lys Leu Lys Ala Leu Val Ala Thr Ala His Ser Glu Leu Ala  
 2585 2590 2595  
 Lys Gly Val Ala Leu Asp Gly Val Leu Ser Thr Phe Val Ser Ala  
 2600 2605 2610

## S226CAS111.ST25

Ala Arg Gln Gly Val Val Asp Thr Asp Val Asp Thr Lys Asp Val  
 2615 2620 2625  
 Ile Glu Cys Leu Lys Leu Ser His His Ser Asp Leu Glu Val Thr  
 2630 2635 2640  
 Gly Asp Ser Cys Asn Asn Phe Met Leu Thr Tyr Asn Lys Val Glu  
 2645 2650 2655  
 Asn Met Thr Pro Arg Asp Leu Gly Ala Cys Ile Asp Cys Asn Ala  
 2660 2665 2670  
 Arg His Ile Asn Ala Gln Val Ala Lys Ser His Asn Val Ser Leu  
 2675 2680 2685  
 Ile Trp Asn Val Lys Asp Tyr Met Ser Leu Ser Glu Gln Leu Arg  
 2690 2695 2700  
 Lys Gln Ile Arg Ser Ala Ala Lys Lys Asn Asn Ile Pro Phe Arg  
 2705 2710 2715  
 Leu Thr Cys Ala Thr Thr Arg Gln Val Val Asn Val Ile Thr Thr  
 2720 2725 2730  
 Lys Ile Ser Leu Lys Gly Gly Lys Ile Val Ser Thr Cys Phe Lys  
 2735 2740 2745  
 Leu Met Leu Lys Ala Thr Leu Leu Cys Val Leu Ala Ala Leu Val  
 2750 2755 2760  
 Cys Tyr Ile Val Met Pro Val His Thr Leu Ser Ile His Asp Gly  
 2765 2770 2775  
 Tyr Thr Asn Glu Ile Ile Gly Tyr Lys Ala Ile Gln Asp Gly Val  
 2780 2785 2790  
 Thr Arg Asp Ile Ile Ser Thr Asp Asp Cys Phe Ala Asn Lys His  
 2795 2800 2805  
 Ala Gly Phe Asp Ala Trp Phe Ser Gln Arg Gly Gly Ser Tyr Lys  
 2810 2815 2820  
 Asn Asp Lys Ser Cys Pro Val Val Ala Ala Ile Ile Thr Arg Glu  
 2825 2830 2835  
 Ile Gly Phe Ile Val Pro Gly Leu Pro Gly Thr Val Leu Arg Ala  
 2840 2845 2850  
 Ile Asn Gly Asp Phe Leu His Phe Leu Pro Arg Val Phe Ser Ala  
 2855 2860 2865

## S226CAS111.ST25

Val Gly Asn Ile Cys Tyr Thr Pro Ser Lys Leu Ile Glu Tyr Ser  
 2870 2875 2880  
 Asp Phe Ala Thr Ser Ala Cys Val Leu Ala Ala Glu Cys Thr Ile  
 2885 2890 2895  
 Phe Lys Asp Ala Met Gly Lys Pro Val Pro Tyr Cys Tyr Asp Thr  
 2900 2905 2910  
 Asn Leu Leu Glu Gly Ser Ile Ser Tyr Ser Glu Leu Arg Pro Asp  
 2915 2920 2925  
 Thr Arg Tyr Val Leu Met Asp Gly Ser Ile Ile Gln Phe Pro Asn  
 2930 2935 2940  
 Thr Tyr Leu Glu Gly Ser Val Arg Val Val Thr Thr Phe Asp Ala  
 2945 2950 2955  
 Glu Tyr Cys Arg His Gly Thr Cys Glu Arg Ser Glu Val Gly Ile  
 2960 2965 2970  
 Cys Leu Ser Thr Ser Gly Arg Trp Val Leu Asn Asn Glu His Tyr  
 2975 2980 2985  
 Arg Ala Leu Ser Gly Val Phe Cys Gly Val Asp Ala Met Asn Leu  
 2990 2995 3000  
 Ile Ala Asn Ile Phe Thr Pro Leu Val Gln Pro Val Gly Ala Leu  
 3005 3010 3015  
 Asp Val Ser Ala Ser Val Val Ala Gly Gly Ile Ile Ala Ile Leu  
 3020 3025 3030  
 Val Thr Cys Ala Ala Tyr Tyr Phe Met Lys Phe Arg Arg Val Phe  
 3035 3040 3045  
 Gly Glu Tyr Asn His Val Val Ala Ala Asn Ala Leu Leu Phe Leu  
 3050 3055 3060  
 Met Ser Phe Thr Ile Leu Cys Leu Val Pro Ala Tyr Ser Phe Leu  
 3065 3070 3075  
 Pro Gly Val Tyr Ser Val Phe Tyr Leu Tyr Leu Thr Phe Tyr Phe  
 3080 3085 3090  
 Thr Asn Asp Val Ser Phe Leu Ala His Leu Gln Trp Phe Ala Met  
 3095 3100 3105  
 Phe Ser Pro Ile Val Pro Phe Trp Ile Thr Ala Ile Tyr Val Phe  
 3110 3115 3120

S226CAS111.ST25

Cys	Ile	Ser	Leu	Lys	His	Cys	His	Trp	Phe	Phe	Asn	Asn	Tyr	Leu
	3125					3130					3135			
Arg	Lys	Arg	Val	Met	Phe	Asn	Gly	Val	Thr	Phe	Ser	Thr	Phe	Glu
	3140					3145					3150			
Glu	Ala	Ala	Leu	Cys	Thr	Phe	Leu	Leu	Asn	Lys	Glu	Met	Tyr	Leu
	3155					3160					3165			
Lys	Leu	Arg	Ser	Glu	Thr	Leu	Leu	Pro	Leu	Thr	Gln	Tyr	Asn	Arg
	3170					3175					3180			
Tyr	Leu	Ala	Leu	Tyr	Asn	Lys	Tyr	Lys	Tyr	Phe	Ser	Gly	Ala	Leu
	3185					3190					3195			
Asp	Thr	Thr	Ser	Tyr	Arg	Glu	Ala	Ala	Cys	Cys	His	Leu	Ala	Lys
	3200					3205					3210			
Ala	Leu	Asn	Asp	Phe	Ser	Asn	Ser	Gly	Ala	Asp	Val	Leu	Tyr	Gln
	3215					3220					3225			
Pro	Pro	Gln	Thr	Ser	Ile	Thr	Ser	Ala	Val	Leu	Gln	Ser	Gly	Phe
	3230					3235					3240			
Arg	Lys	Met	Ala	Phe	Pro	Ser	Gly	Lys	Val	Glu	Gly	Cys	Met	Val
	3245					3250					3255			
Gln	Val	Thr	Cys	Gly	Thr	Thr	Thr	Leu	Asn	Gly	Leu	Trp	Leu	Asp
	3260					3265					3270			
Asp	Thr	Val	Tyr	Cys	Pro	Arg	His	Val	Ile	Cys	Thr	Ala	Glu	Asp
	3275					3280					3285			
Met	Leu	Asn	Pro	Asn	Tyr	Glu	Asp	Leu	Leu	Ile	Arg	Lys	Ser	Asn
	3290					3295					3300			
His	Ser	Phe	Leu	Val	Gln	Ala	Gly	Asn	Val	Gln	Leu	Arg	Val	Ile
	3305					3310					3315			
Gly	His	Ser	Met	Gln	Asn	Cys	Leu	Leu	Arg	Leu	Lys	Val	Asp	Thr
	3320					3325					3330			
Ser	Asn	Pro	Lys	Thr	Pro	Lys	Tyr	Lys	Phe	Val	Arg	Ile	Gln	Pro
	3335					3340					3345			
Gly	Gln	Thr	Phe	Ser	Val	Leu	Ala	Cys	Tyr	Asn	Gly	Ser	Pro	Ser
	3350					3355					3360			
Gly	Val	Tyr	Gln	Cys	Ala	Met	Arg	Pro	Asn	His	Thr	Ile	Lys	Gly
	3365					3370					3375			

## S226CAS111.ST25

Ser Phe Leu Asn Gly Ser Cys Gly Ser Val Gly Phe Asn Ile Asp  
 3380 3385 3390  
 Tyr Asp Cys Val Ser Phe Cys Tyr Met His His Met Glu Leu Pro  
 3395 3400 3405  
 Thr Gly Val His Ala Gly Thr Asp Leu Glu Gly Lys Phe Tyr Gly  
 3410 3415 3420  
 Pro Phe Val Asp Arg Gln Thr Ala Gln Ala Ala Gly Thr Asp Thr  
 3425 3430 3435  
 Thr Ile Thr Leu Asn Val Leu Ala Trp Leu Tyr Ala Ala Val Ile  
 3440 3445 3450  
 Asn Gly Asp Arg Trp Phe Leu Asn Arg Phe Thr Thr Thr Leu Asn  
 3455 3460 3465  
 Asp Phe Asn Leu Val Ala Met Lys Tyr Asn Tyr Glu Pro Leu Thr  
 3470 3475 3480  
 Gln Asp His Val Asp Ile Leu Gly Pro Leu Ser Ala Gln Thr Gly  
 3485 3490 3495  
 Ile Ala Val Leu Asp Met Cys Ala Ala Leu Lys Glu Leu Leu Gln  
 3500 3505 3510  
 Asn Gly Met Asn Gly Arg Thr Ile Leu Gly Ser Thr Ile Leu Glu  
 3515 3520 3525  
 Asp Glu Phe Thr Pro Phe Asp Val Val Arg Gln Cys Ser Gly Val  
 3530 3535 3540  
 Thr Phe Gln Gly Lys Phe Lys Lys Ile Val Lys Gly Thr His His  
 3545 3550 3555  
 Trp Met Leu Leu Thr Phe Leu Thr Ser Leu Leu Ile Leu Val Gln  
 3560 3565 3570  
 Ser Thr Gln Trp Ser Leu Phe Phe Phe Val Tyr Glu Asn Ala Phe  
 3575 3580 3585  
 Leu Pro Phe Thr Leu Gly Ile Met Ala Ile Ala Ala Cys Ala Met  
 3590 3595 3600  
 Leu Leu Val Lys His Lys His Ala Phe Leu Cys Leu Phe Leu Leu  
 3605 3610 3615  
 Pro Ser Leu Ala Thr Val Ala Tyr Phe Asn Met Val Tyr Met Pro  
 3620 3625 3630



## S226CAS111.ST25

Ala Ser Trp Val Met Arg Ile Met Thr Trp Leu Glu Leu Ala Asp  
 3635 3640 3645  
 Thr Ser Leu Ser Gly Tyr Arg Leu Lys Asp Cys Val Met Tyr Ala  
 3650 3655 3660  
 Ser Ala Leu Val Leu Leu Ile Leu Met Thr Ala Arg Thr Val Tyr  
 3665 3670 3675  
 Asp Asp Ala Ala Arg Arg Val Trp Thr Leu Met Asn Val Ile Thr  
 3680 3685 3690  
 Leu Val Tyr Lys Val Tyr Tyr Gly Asn Ala Leu Asp Gln Ala Ile  
 3695 3700 3705  
 Ser Met Trp Ala Leu Val Ile Ser Val Thr Ser Asn Tyr Ser Gly  
 3710 3715 3720  
 Val Val Thr Thr Ile Met Phe Leu Ala Arg Ala Ile Val Phe Val  
 3725 3730 3735  
 Cys Val Glu Tyr Tyr Pro Leu Leu Phe Ile Thr Gly Asn Thr Leu  
 3740 3745 3750  
 Gln Cys Ile Met Leu Val Tyr Cys Phe Leu Gly Tyr Cys Cys Cys  
 3755 3760 3765  
 Cys Tyr Phe Gly Leu Phe Cys Leu Leu Asn Arg Tyr Phe Arg Leu  
 3770 3775 3780  
 Thr Leu Gly Val Tyr Asp Tyr Leu Val Ser Thr Gln Glu Phe Arg  
 3785 3790 3795  
 Tyr Met Asn Ser Gln Gly Leu Leu Pro Pro Lys Ser Ser Ile Asp  
 3800 3805 3810  
 Ala Phe Lys Leu Asn Ile Lys Leu Leu Gly Ile Gly Gly Lys Pro  
 3815 3820 3825  
 Cys Ile Lys Val Ala Thr Val Gln Ser Lys Met Ser Asp Val Lys  
 3830 3835 3840  
 Cys Thr Ser Val Val Leu Leu Ser Val Leu Gln Gln Leu Arg Val  
 3845 3850 3855  
 Glu Ser Ser Ser Lys Leu Trp Ala Gln Cys Val Gln Leu His Asn  
 3860 3865 3870  
 Asp Ile Leu Leu Ala Lys Asp Thr Thr Glu Ala Phe Glu Lys Met  
 3875 3880 3885

## S226CAS111.ST25

Val Ser Leu Leu Ser Val Leu Leu Ser Met Gln Gly Ala Val Asp  
 3890 3895 3900  
 Ile Asn Arg Leu Cys Glu Glu Met Leu Asp Asn Arg Ala Thr Leu  
 3905 3910 3915  
 Gln Ala Ile Ala Ser Glu Phe Ser Ser Leu Pro Ser Tyr Ala Ala  
 3920 3925 3930  
 Tyr Ala Thr Ala Gln Glu Ala Tyr Glu Gln Ala Val Ala Asn Gly  
 3935 3940 3945  
 Asp Ser Glu Val Val Leu Lys Lys Leu Lys Lys Ser Leu Asn Val  
 3950 3955 3960  
 Ala Lys Ser Glu Phe Asp Arg Asp Ala Ala Met Gln Arg Lys Leu  
 3965 3970 3975  
 Glu Lys Met Ala Asp Gln Ala Met Thr Gln Met Tyr Lys Gln Ala  
 3980 3985 3990  
 Arg Ser Glu Asp Lys Arg Ala Lys Val Thr Ser Ala Met Gln Thr  
 3995 4000 4005  
 Met Leu Phe Thr Met Leu Arg Lys Leu Asp Asn Asp Ala Leu Asn  
 4010 4015 4020  
 Asn Ile Ile Asn Asn Ala Arg Asp Gly Cys Val Pro Leu Asn Ile  
 4025 4030 4035  
 Ile Pro Leu Thr Thr Ala Ala Lys Leu Met Val Val Val Pro Asp  
 4040 4045 4050  
 Tyr Gly Thr Tyr Lys Asn Thr Cys Asp Gly Asn Thr Phe Thr Tyr  
 4055 4060 4065  
 Ala Ser Ala Leu Trp Glu Ile Gln Gln Val Val Asp Ala Asp Ser  
 4070 4075 4080  
 Lys Ile Val Gln Leu Ser Glu Ile Asn Met Asp Asn Ser Pro Asn  
 4085 4090 4095  
 Leu Ala Trp Pro Leu Ile Val Thr Ala Leu Arg Ala Asn Ser Ala  
 4100 4105 4110  
 Val Lys Leu Gln Asn Asn Glu Leu Ser Pro Val Ala Leu Arg Gln  
 4115 4120 4125  
 Met Ser Cys Ala Ala Gly Thr Thr Gln Thr Ala Cys Thr Asp Asp  
 4130 4135 4140

S226CAS111.ST25

Asn Ala Leu Ala Tyr Tyr Asn Asn Ser Lys Gly Gly Arg Phe Val  
 4145 4150 4155

Leu Ala Leu Leu Ser Asp His Gln Asp Leu Lys Trp Ala Arg Phe  
 4160 4165 4170

Pro Lys Ser Asp Gly Thr Gly Thr Ile Tyr Thr Glu Leu Glu Pro  
 4175 4180 4185

Pro Cys Arg Phe Val Thr Asp Thr Pro Lys Gly Pro Lys Val Lys  
 4190 4195 4200

Tyr Leu Tyr Phe Ile Lys Gly Leu Asn Asn Leu Asn Arg Gly Met  
 4205 4210 4215

Val Leu Gly Ser Leu Ala Ala Thr Val Arg Leu Gln Ala Gly Asn  
 4220 4225 4230

Ala Thr Glu Val Pro Ala Asn Ser Thr Val Leu Ser Phe Cys Ala  
 4235 4240 4245

Phe Ala Val Asp Pro Ala Lys Ala Tyr Lys Asp Tyr Leu Ala Ser  
 4250 4255 4260

Gly Gly Gln Pro Ile Thr Asn Cys Val Lys Met Leu Cys Thr His  
 4265 4270 4275

Thr Gly Thr Gly Gln Ala Ile Thr Val Thr Pro Glu Ala Asn Met  
 4280 4285 4290

Asp Gln Glu Ser Phe Gly Gly Ala Ser Cys Cys Leu Tyr Cys Arg  
 4295 4300 4305

Cys His Ile Asp His Pro Asn Pro Lys Gly Phe Cys Asp Leu Lys  
 4310 4315 4320

Gly Lys Tyr Val Gln Ile Pro Thr Thr Cys Ala Asn Asp Pro Val  
 4325 4330 4335

Gly Phe Thr Leu Arg Asn Thr Val Cys Thr Val Cys Gly Met Trp  
 4340 4345 4350

Lys Gly Tyr Gly Cys Ser Cys Asp Gln Leu Arg Glu Pro Leu Met  
 4355 4360 4365

Gln Ser Ala Asp Ala Ser Thr Phe Leu Asn Gly Phe Ala Val  
 4370 4375 4380

<210> 75<211> 2695<212> PRT<213> CORONAVIRUS<400> 75

Arg Val Cys Gly Val Ser Ala Ala Arg Leu Thr Pro Cys Gly Thr Gly

1

5

10

15

Thr Ser Thr Asp Val Val Tyr Arg Ala Phe Asp Ile Tyr Asn Glu Lys  
20 25 30

Val Ala Gly Phe Ala Lys Phe Leu Lys Thr Asn Cys Cys Arg Phe Gln  
35 40 45

Glu Lys Asp Glu Glu Gly Asn Leu Leu Asp Ser Tyr Phe Val Val Lys  
50 55 60

Arg His Thr Met Ser Asn Tyr Gln His Glu Glu Thr Ile Tyr Asn Leu  
65 70 75 80

Val Lys Asp Cys Pro Ala Val Ala Val His Asp Phe Phe Lys Phe Arg  
85 90 95

Val Asp Gly Asp Met Val Pro His Ile Ser Arg Gln Arg Leu Thr Lys  
100 105 110

Tyr Thr Met Ala Asp Leu Val Tyr Ala Leu Arg His Phe Asp Glu Gly  
115 120 125

Asn Cys Asp Thr Leu Lys Glu Ile Leu Val Thr Tyr Asn Cys Cys Asp  
130 135 140

Asp Asp Tyr Phe Asn Lys Lys Asp Trp Tyr Asp Phe Val Glu Asn Pro  
145 150 155 160

Asp Ile Leu Arg Val Tyr Ala Asn Leu Gly Glu Arg Val Arg Gln Ser  
165 170 175

Leu Leu Lys Thr Val Gln Phe Cys Asp Ala Met Arg Asp Ala Gly Ile  
180 185 190

Val Gly Val Leu Thr Leu Asp Asn Gln Asp Leu Asn Gly Asn Trp Tyr  
195 200 205

Asp Phe Gly Asp Phe Val Gln Val Ala Pro Gly Cys Gly Val Pro Ile  
210 215 220

Val Asp Ser Tyr Tyr Ser Leu Leu Met Pro Ile Leu Thr Leu Thr Arg  
225 230 235 240

Ala Leu Ala Ala Glu Ser His Met Asp Ala Asp Leu Ala Lys Pro Leu  
245 250 255

Ile Lys Trp Asp Leu Leu Lys Tyr Asp Phe Thr Glu Glu Arg Leu Cys  
260 265 270

Leu Phe Asp Arg Tyr Phe Lys Tyr Trp Asp Gln Thr Tyr His Pro Asn  
Page 93

275

280

285

Cys Ile Asn Cys Leu Asp Asp Arg Cys Ile Leu His Cys Ala Asn Phe  
 290 295 300

Asn Val Leu Phe Ser Thr Val Phe Pro Pro Thr Ser Phe Gly Pro Leu  
 305 310 315 320

Val Arg Lys Ile Phe Val Asp Gly Val Pro Phe Val Val Ser Thr Gly  
 325 330 335

Tyr His Phe Arg Glu Leu Gly Val Val His Asn Gln Asp Val Asn Leu  
 340 345 350

His Ser Ser Arg Leu Ser Phe Lys Glu Leu Leu Val Tyr Ala Ala Asp  
 355 360 365

Pro Ala Met His Ala Ala Ser Gly Asn Leu Leu Leu Asp Lys Arg Thr  
 370 375 380

Thr Cys Phe Ser Val Ala Ala Leu Thr Asn Asn Val Ala Phe Gln Thr  
 385 390 395 400

Val Lys Pro Gly Asn Phe Asn Lys Asp Phe Tyr Asp Phe Ala Val Ser  
 405 410 415

Lys Gly Phe Phe Lys Glu Gly Ser Ser Val Glu Leu Lys His Phe Phe  
 420 425 430

Phe Ala Gln Asp Gly Asn Ala Ala Ile Ser Asp Tyr Asp Tyr Tyr Arg  
 435 440 445

Tyr Asn Leu Pro Thr Met Cys Asp Ile Arg Gln Leu Leu Phe Val Val  
 450 455 460

Glu Val Val Asp Lys Tyr Phe Asp Cys Tyr Asp Gly Gly Cys Ile Asn  
 465 470 475 480

Ala Asn Gln Val Ile Val Asn Asn Leu Asp Lys Ser Ala Gly Phe Pro  
 485 490 495

Phe Asn Lys Trp Gly Lys Ala Arg Leu Tyr Tyr Asp Ser Met Ser Tyr  
 500 505 510

Glu Asp Gln Asp Ala Leu Phe Ala Tyr Thr Lys Arg Asn Val Ile Pro  
 515 520 525

Thr Ile Thr Gln Met Asn Leu Lys Tyr Ala Ile Ser Ala Lys Asn Arg  
 530 535 540

Ala Arg Thr Val Ala Gly Val Ser Ile Cys Ser Thr Met Thr Asn Arg  
 Page 94

## S226CAS111.ST25

545                      550                      555                      560  
 Gln Phe His Gln Lys<sub>565</sub> Leu Leu Lys Ser Ile<sub>570</sub> Ala Ala Thr Arg Gly<sub>575</sub> Ala  
 Thr Val Val Ile<sub>580</sub> Gly Thr Ser Lys Phe<sub>585</sub> Tyr Gly Gly Trp His<sub>590</sub> Asn Met  
 Leu Lys Thr<sub>595</sub> Val Tyr Ser Asp Val<sub>600</sub> Glu Thr Pro His<sub>605</sub> Leu Met Gly Trp  
 Asp Tyr<sub>610</sub> Pro Lys Cys Asp Arg<sub>615</sub> Ala Met Pro Asn Met<sub>620</sub> Leu Arg Ile Met  
 Ala<sub>625</sub> Ser Leu Val Leu Ala<sub>630</sub> Arg Lys His Asn Thr<sub>635</sub> Cys Cys Asn Leu Ser<sub>640</sub>  
 His Arg Phe Tyr Arg<sub>645</sub> Leu Ala Asn Glu Cys<sub>650</sub> Ala Gln Val Leu Ser<sub>655</sub> Glu  
 Met Val Met Cys<sub>660</sub> Gly Gly Ser Leu Tyr<sub>665</sub> Val Lys Pro Gly Gly<sub>670</sub> Thr Ser  
 Ser Gly Asp<sub>675</sub> Ala Thr Thr Ala Tyr<sub>680</sub> Ala Asn Ser Val Phe<sub>685</sub> Asn Ile Cys  
 Gln Ala<sub>690</sub> Val Thr Ala Asn Val<sub>695</sub> Asn Ala Leu Leu Ser<sub>700</sub> Thr Asp Gly Asn  
 Lys<sub>705</sub> Ile Ala Asp Lys Tyr<sub>710</sub> Val Arg Asn Leu Gln<sub>715</sub> His Arg Leu Tyr Glu<sub>720</sub>  
 Cys Leu Tyr Arg Asn<sub>725</sub> Arg Asp Val Asp<sub>730</sub> His Glu Phe Val Asp<sub>735</sub> Glu Phe  
 Tyr Ala Tyr Leu<sub>740</sub> Arg Lys His Phe Ser<sub>745</sub> Met Met Ile Leu Ser<sub>750</sub> Asp Asp  
 Ala Val Val<sub>755</sub> Cys Tyr Asn Ser Asn<sub>760</sub> Tyr Ala Ala Gln Gly<sub>765</sub> Leu Val Ala  
 Ser Ile<sub>770</sub> Lys Asn Phe Lys Ala<sub>775</sub> Val Leu Tyr Tyr Gln<sub>780</sub> Asn Asn Val Phe  
 Met<sub>785</sub> Ser Glu Ala Lys Cys<sub>790</sub> Trp Thr Glu Thr Asp<sub>795</sub> Leu Thr Lys Gly Pro<sub>800</sub>  
 His Glu Phe Cys Ser<sub>805</sub> Gln His Thr Met Leu<sub>810</sub> Val Lys Gln Gly Asp<sub>815</sub> Asp  
 Tyr Val Tyr Leu Pro Tyr Pro Asp Pro Ser Arg Ile Leu Gly Ala Gly  
 Page 95

Cys Phe Val Asp Asp Ile Val Lys Thr Asp Gly Thr Leu Met Ile Glu  
835 840 845

Arg Phe Val Ser Leu Ala Ile Asp Ala Tyr Pro Leu Thr Lys His Pro  
850 855 860

Asn Gln Glu Tyr Ala Asp Val Phe His Leu Tyr Leu Gln Tyr Ile Arg  
865 870 875 880

Lys Leu His Asp Glu Leu Thr Gly His Met Leu Asp Met Tyr Ser Val  
885 890 895

Met Leu Thr Asn Asp Asn Thr Ser Arg Tyr Trp Glu Pro Glu Phe Tyr  
900 905 910

Glu Ala Met Tyr Thr Pro His Thr Val Leu Gln Ala Val Gly Ala Cys  
915 920 925

Val Leu Cys Asn Ser Gln Thr Ser Leu Arg Cys Gly Ala Cys Ile Arg  
930 935 940

Arg Pro Phe Leu Cys Cys Lys Cys Cys Tyr Asp His Val Ile Ser Thr  
945 950 955 960

Ser His Lys Leu Val Leu Ser Val Asn Pro Tyr Val Cys Asn Ala Pro  
965 970 975

Gly Cys Asp Val Thr Asp Val Thr Gln Leu Tyr Leu Gly Gly Met Ser  
980 985 990

Tyr Tyr Cys Lys Ser His Lys Pro Pro Ile Ser Phe Pro Leu Cys Ala  
995 1000 1005

Asn Gly Gln Val Phe Gly Leu Tyr Lys Asn Thr Cys Val Gly Ser  
1010 1015 1020

Asp Asn Val Thr Asp Phe Asn Ala Ile Ala Thr Cys Asp Trp Thr  
1025 1030 1035

Asn Ala Gly Asp Tyr Ile Leu Ala Asn Thr Cys Thr Glu Arg Leu  
1040 1045 1050

Lys Leu Phe Ala Ala Glu Thr Leu Lys Ala Thr Glu Glu Thr Phe  
1055 1060 1065

Lys Leu Ser Tyr Gly Ile Ala Thr Val Arg Glu Val Leu Ser Asp  
1070 1075 1080

Arg Glu Leu His Leu Ser Trp Glu Val Gly Lys Pro Arg Pro Pro

## S226CAS111.ST25

1085

1090

1095

Leu	Asn	Arg	Asn	Tyr	Val	Phe	Thr	Gly	Tyr	Arg	Val	Thr	Lys	Asn
1100						1105					1110			
Ser	Lys	Val	Gln	Ile	Gly	Glu	Tyr	Thr	Phe	Glu	Lys	Gly	Asp	Tyr
1115						1120					1125			
Gly	Asp	Ala	Val	Val	Tyr	Arg	Gly	Thr	Thr	Thr	Tyr	Lys	Leu	Asn
1130						1135					1140			
Val	Gly	Asp	Tyr	Phe	Val	Leu	Thr	Ser	His	Thr	Val	Met	Pro	Leu
1145						1150					1155			
Ser	Ala	Pro	Thr	Leu	Val	Pro	Gln	Glu	His	Tyr	Val	Arg	Ile	Thr
1160						1165					1170			
Gly	Leu	Tyr	Pro	Thr	Leu	Asn	Ile	Ser	Asp	Glu	Phe	Ser	Ser	Asn
1175						1180					1185			
Val	Ala	Asn	Tyr	Gln	Lys	Val	Gly	Met	Gln	Lys	Tyr	Ser	Thr	Leu
1190						1195					1200			
Gln	Gly	Pro	Pro	Gly	Thr	Gly	Lys	Ser	His	Phe	Ala	Ile	Gly	Leu
1205						1210					1215			
Ala	Leu	Tyr	Tyr	Pro	Ser	Ala	Arg	Ile	Val	Tyr	Thr	Ala	Cys	Ser
1220						1225					1230			
His	Ala	Ala	Val	Asp	Ala	Leu	Cys	Glu	Lys	Ala	Leu	Lys	Tyr	Leu
1235						1240					1245			
Pro	Ile	Asp	Lys	Cys	Ser	Arg	Ile	Ile	Pro	Ala	Arg	Ala	Arg	Val
1250						1255					1260			
Glu	Cys	Phe	Asp	Lys	Phe	Lys	Val	Asn	Ser	Thr	Leu	Glu	Gln	Tyr
1265						1270					1275			
Val	Phe	Cys	Thr	Val	Asn	Ala	Leu	Pro	Glu	Thr	Thr	Ala	Asp	Ile
1280						1285					1290			
Val	Val	Phe	Asp	Glu	Ile	Ser	Met	Ala	Thr	Asn	Tyr	Asp	Leu	Ser
1295						1300					1305			
Val	Val	Asn	Ala	Arg	Leu	Arg	Ala	Lys	His	Tyr	Val	Tyr	Ile	Gly
1310						1315					1320			
Asp	Pro	Ala	Gln	Leu	Pro	Ala	Pro	Arg	Thr	Leu	Leu	Thr	Lys	Gly
1325						1330					1335			
Thr	Leu	Glu	Pro	Glu	Tyr	Phe	Asn	Ser	Val	Cys	Arg	Leu	Met	Lys



1340 1345 1350  
 Thr Ile Gly Pro Asp Met Phe Leu Gly Thr Cys Arg Arg Cys Pro  
 1355 1360 1365  
 Ala Glu Ile Val Asp Thr Val Ser Ala Leu Val Tyr Asp Asn Lys  
 1370 1375 1380  
 Leu Lys Ala His Lys Asp Lys Ser Ala Gln Cys Phe Lys Met Phe  
 1385 1390 1395  
 Tyr Lys Gly Val Ile Thr His Asp Val Ser Ser Ala Ile Asn Arg  
 1400 1405 1410  
 Pro Gln Ile Gly Val Val Arg Glu Phe Leu Thr Arg Asn Pro Ala  
 1415 1420 1425  
 Trp Arg Lys Ala Val Phe Ile Ser Pro Tyr Asn Ser Gln Asn Ala  
 1430 1435 1440  
 Val Ala Ser Lys Ile Leu Gly Leu Pro Thr Gln Thr Val Asp Ser  
 1445 1450 1455  
 Ser Gln Gly Ser Glu Tyr Asp Tyr Val Ile Phe Thr Gln Thr Thr  
 1460 1465 1470  
 Glu Thr Ala His Ser Cys Asn Val Asn Arg Phe Asn Val Ala Ile  
 1475 1480 1485  
 Thr Arg Ala Lys Ile Gly Ile Leu Cys Ile Met Ser Asp Arg Asp  
 1490 1495 1500  
 Leu Tyr Asp Lys Leu Gln Phe Thr Ser Leu Glu Ile Pro Arg Arg  
 1505 1510 1515  
 Asn Val Ala Thr Leu Gln Ala Glu Asn Val Thr Gly Leu Phe Lys  
 1520 1525 1530  
 Asp Cys Ser Lys Ile Ile Thr Gly Leu His Pro Thr Gln Ala Pro  
 1535 1540 1545  
 Thr His Leu Ser Val Asp Ile Lys Phe Lys Thr Glu Gly Leu Cys  
 1550 1555 1560  
 Val Asp Ile Pro Gly Ile Pro Lys Asp Met Thr Tyr Arg Arg Leu  
 1565 1570 1575  
 Ile Ser Met Met Gly Phe Lys Met Asn Tyr Gln Val Asn Gly Tyr  
 1580 1585 1590  
 Pro Asn Met Phe Ile Thr Arg Glu Glu Ala Ile Arg His Val Arg

## S226CAS111.ST25

1595

1600

1605

Ala Trp Ile Gly Phe Asp Val Glu Gly Cys His Ala Thr Arg Asp  
1610 1615 1620

Ala Val Gly Thr Asn Leu Pro Leu Gln Leu Gly Phe Ser Thr Gly  
1625 1630 1635

Val Asn Leu Val Ala Val Pro Thr Gly Tyr Val Asp Thr Glu Asn  
1640 1645 1650

Asn Thr Glu Phe Thr Arg Val Asn Ala Lys Pro Pro Pro Gly Asp  
1655 1660 1665

Gln Phe Lys His Leu Ile Pro Leu Met Tyr Lys Gly Leu Pro Trp  
1670 1675 1680

Asn Val Val Arg Ile Lys Ile Val Gln Met Leu Ser Asp Thr Leu  
1685 1690 1695

Lys Gly Leu Ser Asp Arg Val Val Phe Val Leu Trp Ala His Gly  
1700 1705 1710

Phe Glu Leu Thr Ser Met Lys Tyr Phe Val Lys Ile Gly Pro Glu  
1715 1720 1725

Arg Thr Cys Cys Leu Cys Asp Lys Arg Ala Thr Cys Phe Ser Thr  
1730 1735 1740

Ser Ser Asp Thr Tyr Ala Cys Trp Asn His Ser Val Gly Phe Asp  
1745 1750 1755

Tyr Val Tyr Asn Pro Phe Met Ile Asp Val Gln Gln Trp Gly Phe  
1760 1765 1770

Thr Gly Asn Leu Gln Ser Asn His Asp Gln His Cys Gln Val His  
1775 1780 1785

Gly Asn Ala His Val Ala Ser Cys Asp Ala Ile Met Thr Arg Cys  
1790 1795 1800

Leu Ala Val His Glu Cys Phe Val Lys Arg Val Asp Trp Ser Val  
1805 1810 1815

Glu Tyr Pro Ile Ile Gly Asp Glu Leu Arg Val Asn Ser Ala Cys  
1820 1825 1830

Arg Lys Val Gln His Met Val Val Lys Ser Ala Leu Leu Ala Asp  
1835 1840 1845

Lys Phe Pro Val Leu His Asp Ile Gly Asn Pro Lys Ala Ile Lys

1850		1855		1860	
Cys Val	Pro Gln Ala Glu	Val Glu Trp Lys Phe	Tyr Asp Ala Gln		
1865		1870	1875		
Pro Cys	Ser Asp Lys Ala	Tyr Lys Ile Glu Glu	Leu Phe Tyr Ser		
1880		1885	1890		
Tyr Ala	Thr His His Asp	Lys Phe Thr Asp Gly	Val Cys Leu Phe		
1895		1900	1905		
Trp Asn	Cys Asn Val Asp	Arg Tyr Pro Ala Asn	Ala Ile Val Cys		
1910		1915	1920		
Arg Phe	Asp Thr Arg Val	Leu Ser Asn Leu Asn	Leu Pro Gly Cys		
1925		1930	1935		
Asp Gly	Gly Ser Leu Tyr	Val Asn Lys His Ala	Phe His Thr Pro		
1940		1945	1950		
Ala Phe	Asp Lys Ser Ala	Phe Thr Asn Leu Lys	Gln Leu Pro Phe		
1955		1960	1965		
Phe Tyr	Tyr Ser Asp Ser	Pro Cys Glu Ser His	Gly Lys Gln Val		
1970		1975	1980		
Val Ser	Asp Ile Asp Tyr	Val Pro Leu Lys Ser	Ala Thr Cys Ile		
1985		1990	1995		
Thr Arg	Cys Asn Leu Gly	Gly Ala Val Cys Arg	His His Ala Asn		
2000		2005	2010		
Glu Tyr	Arg Gln Tyr Leu	Asp Ala Tyr Asn Met	Met Ile Ser Ala		
2015		2020	2025		
Gly Phe	Ser Leu Trp Ile	Tyr Lys Gln Phe Asp	Thr Tyr Asn Leu		
2030		2035	2040		
Trp Asn	Thr Phe Thr Arg	Leu Gln Ser Leu Glu	Asn Val Ala Tyr		
2045		2050	2055		
Asn Val	Val Asn Lys Gly	His Phe Asp Gly His	Ala Gly Glu Ala		
2060		2065	2070		
Pro Val	Ser Ile Ile Asn	Asn Ala Val Tyr Thr	Lys Val Asp Gly		
2075		2080	2085		
Ile Asp	Val Glu Ile Phe	Glu Asn Lys Thr Thr	Leu Pro Val Asn		
2090		2095	2100		
Val Ala	Phe Glu Leu Trp	Ala Lys Arg Asn Ile	Lys Pro Val Pro		

## S226CAS111.ST25

2105

2110

2115

Glu Ile 2120	Lys Ile	Leu Asn	Asn 2125	Leu Gly	Val Asp	Ile 2130	Ala Ala	Asn
Thr Val 2135	Ile Trp	Asp Tyr	Lys 2140	Arg Glu	Ala Pro	Ala 2145	His Val	Ser
Thr Ile 2150	Gly Val	Cys Thr	Met 2155	Thr Asp	Ile Ala	Lys 2160	Lys Pro	Thr
Glu Ser 2165	Ala Cys	Ser Ser	Leu 2170	Thr Val	Leu Phe	Asp 2175	Gly Arg	Val
Glu Gly 2180	Gln Val	Asp Leu	Phe 2185	Arg Asn	Ala Arg	Asn 2190	Gly Val	Leu
Ile Thr 2195	Glu Gly	Ser Val	Lys 2200	Gly Leu	Thr Pro	Ser 2205	Lys Gly	Pro
Ala Gln 2210	Ala Ser	Val Asn	Gly 2215	Val Thr	Leu Ile	Gly 2220	Glu Ser	Val
Lys Thr 2225	Gln Phe	Asn Tyr	Phe 2230	Lys Lys	Val Asp	Gly 2235	Ile Ile	Gln
Gln Leu 2240	Pro Glu	Thr Tyr	Phe 2245	Thr Gln	Ser Arg	Asp 2250	Leu Glu	Asp
Phe Lys 2255	Pro Arg	Ser Gln	Met 2260	Glu Thr	Asp Phe	Leu 2265	Glu Leu	Ala
Met Asp 2270	Glu Phe	Ile Gln	Arg 2275	Tyr Lys	Leu Glu	Gly 2280	Tyr Ala	Phe
Glu His 2285	Ile Val	Tyr Gly	Asp 2290	Phe Ser	His Gly	Gln 2295	Leu Gly	Gly
Leu His 2300	Leu Met	Ile Gly	Leu 2305	Ala Lys	Arg Ser	Gln 2310	Asp Ser	Pro
Leu Lys 2315	Leu Glu	Asp Phe	Ile 2320	Pro Met	Asp Ser	Thr 2325	Val Lys	Asn
Tyr Phe 2330	Ile Thr	Asp Ala	Gln 2335	Thr Gly	Ser Ser	Lys 2340	Cys Val	Cys
Ser Val 2345	Ile Asp	Leu Leu	Leu 2350	Asp Asp	Phe Val	Glu 2355	Ile Ile	Lys
Ser Gln	Asp Leu	Ser Val	Ile	Ser Lys	Val Val	Lys	Val Thr	Ile

S226CAS111.ST25  
2370

2360

2365

2370

Asp Tyr Ala Glu Ile Ser Phe Met Leu Trp Cys Lys Asp Gly His  
2375 2380 2385

Val Glu Thr Phe Tyr Pro Lys Leu Gln Ala Ser Gln Ala Trp Gln  
2390 2395 2400

Pro Gly Val Ala Met Pro Asn Leu Tyr Lys Met Gln Arg Met Leu  
2405 2410 2415

Leu Glu Lys Cys Asp Leu Gln Asn Tyr Gly Glu Asn Ala Val Ile  
2420 2425 2430

Pro Lys Gly Ile Met Met Asn Val Ala Lys Tyr Thr Gln Leu Cys  
2435 2440 2445

Gln Tyr Leu Asn Thr Leu Thr Leu Ala Val Pro Tyr Asn Met Arg  
2450 2455 2460

Val Ile His Phe Gly Ala Gly Ser Asp Lys Gly Val Ala Pro Gly  
2465 2470 2475

Thr Ala Val Leu Arg Gln Trp Leu Pro Thr Gly Thr Leu Leu Val  
2480 2485 2490

Asp Ser Asp Leu Asn Asp Phe Val Ser Asp Ala Asp Ser Thr Leu  
2495 2500 2505

Ile Gly Asp Cys Ala Thr Val His Thr Ala Asn Lys Trp Asp Leu  
2510 2515 2520

Ile Ile Ser Asp Met Tyr Asp Pro Arg Thr Lys His Val Thr Lys  
2525 2530 2535

Glu Asn Asp Ser Lys Glu Gly Phe Phe Thr Tyr Leu Cys Gly Phe  
2540 2545 2550

Ile Lys Gln Lys Leu Ala Leu Gly Gly Ser Ile Ala Val Lys Ile  
2555 2560 2565

Thr Glu His Ser Trp Asn Ala Asp Leu Tyr Lys Leu Met Gly His  
2570 2575 2580

Phe Ser Trp Trp Thr Ala Phe Val Thr Asn Val Asn Ala Ser Ser  
2585 2590 2595

Ser Glu Ala Phe Leu Ile Gly Ala Asn Tyr Leu Gly Lys Pro Lys  
2600 2605 2610

Glu Gln Ile Asp Gly Tyr Thr Met His Ala Asn Tyr Ile Phe Trp

S226CAS111.ST25

2615

2620

2625

Arg Asn Thr Asn Pro Ile Gln Leu Ser Ser Tyr Ser Leu Phe Asp  
2630 2635 2640

Met Ser Lys Phe Pro Leu Lys Leu Arg Gly Thr Ala Val Met Ser  
2645 2650 2655

Leu Lys Glu Asn Gln Ile Asn Asp Met Ile Tyr Ser Leu Leu Glu  
2660 2665 2670

Lys Gly Arg Leu Ile Ile Arg Glu Asn Asn Arg Val Val Val Ser  
2675 2680 2685

Ser Asp Ile Leu Val Asn Asn  
2690 2695

**FCT/FR 004/003105**



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

☒ **BLACK BORDERS**

☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**

☐ **FADED TEXT OR DRAWING**

☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**

☐ **SKEWED/SLANTED IMAGES**

☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**

☐ **GRAY SCALE DOCUMENTS**

☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**

☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**

☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**